AD_____

Award Number: DAMD17-94-J-4129

TITLE: Identification of BRCA1 and 2 Other Tumor Suppressor Genes on Chromosome 17 Through Positional Cloning

PRINCIPAL INVESTIGATOR: Raymond L. White, Ph.D.

CONTRACTING ORGANIZATION: University of Utah
Salt Lake City, Utah 84102

REPORT DATE: April 2000

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	April 2000	Final (1 Jul 94 - 17 Mar 00)

4. TITLE AND SUBTITLE
Identification of BRCA1 and 2 Other Tumor Suppressor Genes on Chromosome 17 Through Positional Cloning

5. FUNDING NUMBERS
DAMD17-94-J-4129

6. AUTHOR(S)
Raymond L. White, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Utah
Salt Lake City, Utah 84102

E-MAIL:
ray.white@hci.utah.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT ***(Maximum 200 Words)***

The overall goal of this project is to identify genes involved with the development and progression of breast cancer. This goal has remained unchanged since the start of the project, however the discovery of BRCA1 in 1994 together with technological advances in gene expression profiling has influenced our strategy to achieve this goal. In the early part of the project our search for tumor suppressor genes was directed by genetic or LOH mapping strategies followed by positional cloning of candidate genes. As we proposed in our revised statement of work (SOW), we have now focused our efforts entirely on microarray-based comparisons to identify breast cancer related genes. Our laboratory has gained access to this technology through collaboration with Molecular Dynamics and now we have established microarray spotting and scanning systems with over 40,000 minimally redundant sequence-verified human cDNA clones. Once candidate genes have been identified, they will be further characterized using a retroviral-based conditional expression system to assess changes in characteristics pertaining to morphology and growth rate by culturing human breast epithelial cells in an extracellular matrix.

14. SUBJECT TERMS
Tumor Suppressor Genes, LOH Mapping, Chromosome 17, Physical Mapping, Genetic Mapping, CDNA Screening, Humans, Anatomical Samples, Mutation Detection, Breast Cancer

15. NUMBER OF PAGES
81

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

___ Where copyrighted material is quoted, permission has been obtained to use such material.

___ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

___ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

__ In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and use of Laboratory Animals of the Institute of Laboratory Resources, national Research Council (NIH Publication No. 86-23, Revised 1985).

X For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

X In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

X In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

__ In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

Ray White 6/30/00

PI - Signature Date

(4) TABLE OF CONTENTS

(5) INTRODUCTION

The identification and characterization of genes involved in development and progression of breast cancer is critical to an understanding of the biological mechanisms that regulate growth of cells in breast epithelium. The emphasis of this project has been originally directed towards the identification and characterization of several breast cancer tumor suppressor genes located on chromosome 17. Common for each of these genes was that their approximate physical location had been delineated either through genetic linkage analysis of hereditary segregation patterns or by being the target of recurring deleterious events observed as loss of heterozygosity (LOH) in sporadic tumors. Since the discovery of the BRCA1 gene by other groups, our experimental focus have been directed towards the identification of four presumed tumor suppressor genes defined by LOH on chromosome 17, as proposed in our revised statement of work (SOW). Of these four loci, two are located at each extremity of the chromosome, while the remaining two areas of LOH we have identified flank either side of the BRCA1 locus at 17q21. However, contrary to hereditary tumor suppressor genes whose position may be determined by genetic linkage analysis, the localization of tumor suppressor genes by LOH is circumstantial. LOH does not provide direct evidence as to whether the region of LOH actually contains a gene critical to carcinogenesis or whether the deletion is coincidental. Statistically significant areas of LOH indicative of a tumor suppressor gene can only be determined by analyzing large numbers of specimens. At the same time it holds true that coincidental deletions provide grounds for mis-interpretation of the position of an actual tumor suppressor gene. Despite those drawbacks, when our proposal was submitted, this paradigm represented the most efficient approach to identifying tumor suppressor genes. However, it has become exceedingly clear that the massive resources applied in various genome centers to establish both physical and expression maps covering the entire human genome far exceeds the amount of data that we were able to generate even within the modest areas of chromosome 17 where we have concentrated our attention. The past three years have been a period of transition for the project and early on during this period we evaluated the efficiency of our strategy compared to newer, more global strategies to identify genes critically involved with tumor development and progression. We have for a long time realized that LOH guided search for tumor suppressor genes is vulnerable to critique on several levels. Most importantly among these the rather poor delineation of the tumor suppressor loci, which consequently requires a significant expansion of the physical mapping component of the project and thus dramatically augment the number of candidate genes to evaluate. A different and much broader strategy for identifying *any* kind of gene involved in tumor formation and progression, which is not dependent on prior knowledge of location, employs genetic profiling using high density microarrays. Not only does this technique permit detection of the expression profiles of each of a large number of genes in parallel, but it potentially also provides a mechanistic view of how regulatory pathways are controlled. As we proposed in our newly revised statement of work (SOW), we have now focused our efforts entirely on microarray-based comparisons to identify breast cancer related genes. This includes development of robust fluorescent labeling and hybridization protocols as well as the preparation and testing of redundant human cDNA target samples for deposition on the microarray slides. Our laboratory has gained access to this technology through collaboration with Molecular Dynamics who has provided us with an array robot and a scanning device. In addition, during the past year, we have gained over 40,000 minimally redundant sequence-verified human cDNAs for deposition on the microarray slides. As proposed, we have compared expression profiles from several distinct breast cell lines. By scanning 2400 clones on our first slide array we have discovered 29 genes and ESTs that reveal altered expression patterns as a

consequence of conditionally expressed dominant-negative β-catenin in primary breast cells. Once the full-length sequence of each of the differentially regulated genes has been ascertained and prioritized, they will be introduced into appropriate cell lines to study their effect on cell morphology and gene regulation. Experiments have also been carried out to observe regulatory affects of experimentally transformed breast cells grown in matrigel and comparisons of a larger collection of primary cells and established cell-lines is being designed.

(6) BODY

The following progress were made during the funded years

Statement of Work - revised October 23, 1995

Task 1, Isolation and characterization of two potential tumor suppressor genes approximately 1Mb proximal and 1Mb distal of BRCA1.

During the first twelve months of this fundind period, several discoveries have been made with respect to breast cancer. Most significant was the discovery of Miki et al. (1994) and Futreal et al. (1994) of the BRCA1 gene itself; however, the genomic environment of BRCA1 has also revealed several interesting features. Dr. Solomon's group showed that the 5' of BRCA1 gene is in very close proximity to the 5' of the 1A1-3B gene (Brown et al., 1995), raising the possibility of studying transcriptional control (transcriptional dominance, positive or negative interference) among these two genes. Our group has found that the L21 riboprotein is located within the BRCA1 locus, in an intron flanking BRCA1 exon 14. We have made other potentially very interesting observations by analyzing primary prostate cancers for allelic loss; those results have suggested the presence of two additional tumor suppressor genes in the immediate vicinity of BRCA1 (Brothman et al. 1995, see attached reprint). This new information placed us in a position where the resources available through this grant were adequate to pursue to identify the two potential tumor suppressor genes immediately flanking BRCA1 and located within our existing physical map of the BRCA1 region. We have identified and published possible candidate genes from these regions during this period: two novel members of the DLG family of genes, DLG2 (Mazoyer et al. 1995, see attached reprint), and DLG3 (Smith et al. 1996, see attached reprint) and an ADP-ribosylation factor (Smith et al. 1995, see attached reprint).

Common to all members of the Drosophila discs-large family of genes are three distinct structural domains. At the N-terminal are 1-3 somewhat degenerate Drosophila homology regions (DHRs). DHR motifs are approximately 90 amino acids long and have been shown in vitro to bind cytoskeletal proteins of the band 4.1 family. A region with homology to src oncogene motif 3 (SH3) is found in the central part of the discs-large proteins. This motif, approximately 60 amino acids in length, is known to be a site of protein-protein interactions. Finally, a guanylate kinase domain (GK) is found at the C-termini of the Dlg proteins. The function of this domain is the catalytic transfer of phosphate from ATP to GMP, forming GDP. In Drosophila, Dlg has been shown to be a tumor suppressor (Woods et al., 1989). Additional, and quite compelling, evidence for the biological importance of the DLG family of genes was recently discovered when it was found that APC, the tumor suppressor gene responsible for adenomatous polyposis coli (Groden *et al.*, 1991), interacts on the protein level with the human DLG homolog (Matsumine *et al.*, 1996).

ADP-ribosylation factor (ARF4L) is, based on its predicted protein structure, believed to be involved in membrane trafficking and protein secretion. Six protein domains have been identified,

three of which are involved with phosphate/magnesium binding, while the remaining three are involved with guanine nucelotide binding.

In an experiment designed to directly determine if DLG2, DLG3 and ARF4L are the targets for the LOH observed in sporadic breast cancers, we obtained paired tumor and normal tissue samples from 10 individuals who underwent surgical removal of malignant carcinomas. We extracted DNA and RNA from these samples, and subsequently tested the paired DNA samples for loss of heterozygosity (LOH). As indicated in figure 1, we used five highly polymorphic genetic markers located along the long arm of chromosome 17 (Albertsen et al. 1994a, see attached reprint), three of which lie relatively close to the BRCA1 locus 17 (Albertsen et al. 1994b, see attached reprint). From among the ten tumors we identified 3 which displayed LOH around BRCA1. According to the established model for LOH involving tumor suppressor genes, the allele remaining in the tumor sample would harbor the deleterious mutation. Using RNA extracted from these tumors we prepared first-strand cDNAs specific to each of the three genes and submitted these templates for automated sequencing on an ABI373A sequencer (Applied Biosystems, Foster City, CA). As none of the samples we have sequenced have revealed any mutations, we have no evidence so far to indicate that either DLG2, DLG3 or ARF4L is the primary target in the region of LOH flanking BRCA1.

Another part of our project that has yielded good progress during this period is the identification of two novel genes from the region of LOH encompassing the plakoglobin locus. We discovered these genes in collaboration with Dr. Robert Callahan at the National Cancer Institute using the P1 clones 50H1 and 122F4 identified in our laboratory (Albertsen et al. 1994b, see attached reprint). The first of these genes is the presumed human homologue of the mouse FKBP65 gene (Coss *et al.*, 1995). Genes of the FKBP family derive their names from the immunosuppressant macrolide antibiotic FK506, because they mediate its activity (in part) by binding to a ubiquitous family of highly conserved intracellular receptors termed immunophilins (Sigal and Dumont, 1992). Although FK506 is known to block various signal transduction pathways in normal T-cells, FKBP genes (including FKBP65) are expressed in most tissues that have been analyzed. The biological relevance of human FKBP65 to cancer development remains unclear, but once its full-length sequence is ascertained we will undertake a detailed analysis of the gene and its functional domains. The second gene we identified in the plakoglobin region was ascertained by coincidence. As part of the process of determining the genomic structure of the FKBP65 gene, we sequenced several genomic subclones derived from P1 phage clones 50H1 and 122F4. While analyzing the sequence of one of these subclones, named 1H2M, I identified a small collection of human ESTs that shared a segment of almost 300 nucleotides of perfect homology to 1H2M. Further analysis and database comparisons extended the DNA sequence to approximately 1300 nucleotides, and revealed that the novel gene shared homology with no other currently known vertebrate gene. However, the protein translation of the nucleotide sequence showed 40% homology, over a segment of almost 200 amino acids, to an uncharacterized gene from C. elegans. It is impossible at present to predict the biological relevance of the novel gene with respect to tumor formation, but the high degree of protein homology preserved across such distant species suggests a fundamental and probably critical role.

Another gene that occupied a significant amount of our time and effort was DOC-2, whose identification and characterization was published (Albertsen et al. 1996, see attached reprint). A 788-basepair segment of DOC-2 was originally identified by differential display between ovarian carcinoma and normal ovarian epithelial tissue; its expression was greatly reduced or entirely absent

in a panel of 10 ovarian tumors (Mok *et al.*, 1994). Our ascertainment of DOC-2 was based on cDNA screening of a fetal retina library (Stratagene # 937202) with P1 clone 124D3. One of the cDNA clones we identified, 1RA1, harbored a large segment of the genuine DOC-2 gene; however, we did not realize immediately that the 1RA1 cDNA clone was a chimera between DOC-2 and DLG3. Consequently, our original attempt to verify the chromosomal location of the 1RA1 clone clearly indicated that the clone was located in the BRCA1 region. It was not until we were in the process of determining the genomic structure of DOC-2 that we realized our mistake and found the correct genomic location of DOC-2 on chromosome 5. Nevertheless, the complete sequence, genomic characterization, and chromosomal location of DOC-2 have been attributed to the present Army grant.

Additional expressed sequences from the two regions in question exist, and we intend to do this by screening for mutations in these genes in sporadic tumors which reveal LOH. Alternatively, the genes will be tested for their biological roles in tissue culture systems, either through DNA oligo antisense-mediated suppression of normal gene expression or by transfecting breast epithelial cells with conditionally induceible expression vectors carrying the genes.

Statement of Work - revised October 23, 1995

Task 2, To identify and characterize the breast cancer tumor suppressor genes on distal 17q and distal 17p using physical reagents identified by the CEPH.

We initiated a physical mapping project to refine the rather crude physical map of 17p. We have chosen to focus solely on this region in the initial stages of the physical mapping project for two reasons: a) the 17p region of LOH is better characterized and thereby provides a better possibility for identifying the tumor suppressor gene(s) located there; and b) with only limited manpower available to satisfy all aspects of our research proposal it would be unwise to further dilute our efforts by simultaneously attempting to refine the physical map of the 17q region. During the first year two developments relating to the chromosome 17p region led us to initiate a refined physical mapping of 17pter. Of greatest importance was the publication of two novel candidate tumor suppressor genes in a quite narrowly delimited region of 17p13.3 (Schultz *et al.*, 1996). While it is not clear whether either of these genes is the target for LOH near the telomere of the short arm of chromosome 17, an important conclusion can be drawn: the very distal location of the presumed tumor suppressor locus at the extremity of the chromosome means that the general numeric relation between genetic distance measured in cM to the physical length measured in Mb, which exists in the central parts of chromosomes, can not be applied in this case because the frequency of genetic recombinations is elevated in telomeric regions. This phenomenon dramatically skews the relationship between genetic distance and physical distance, such that a large genetic distance actually is contained within a relatively short physical segment of DNA. Schultz *et al.*, (1996) showed that the 3.5 cM of genetic distance between D17S5 and D17S28, which normally would be expected to represent a 3.5-Mb genomic region, in reality is contained within a single cosmid (30kb). If the genetic compression observed in this short telomeric region of the short arm of chromosome 17 extends beyond the confined segment delimited by the markers D17S5 and D17S28, the task of refining the physical map in the approximately 15-cM region which harbors the presumed tumor suppressor gene(s) should be relatively simple. To this end we are now in the process of obtaining yeast artificial chromosomes (YACs) that have been localized to telomeric 17p through the genome mapping efforts at the CEPH and the Whitehead Institute at MIT. By cross referencing these YACs with the genetic markers we have developed and mapped to this region

(Gerken et al. 1995, see attached reprint), we can identify the exact extent of the existing physical coverage of the region. The information obtained in that way will allow us to localize potential gaps in the physical maps and will provide guidance as to where the genomic coverage must be expanded to complete the physical map of the region.

Our initial strategy for identifying tumor suppressor genes in areas of LOH on chromosome 17 was based on a positional cloning strategy. This strategy is divided into three separate stages. During the first stage a physical map of a the area of interest is constructed using genomic clones like YACs, BACs or P1s. Following this stage expressed sequences are ascertained, and as the last stage the candidate genes are analyzed for mutations in breast tumors to determine if indeed they are the sought-for tumor suppressor gene. While we successfully have used this approach in the past (Cawthon et al. 1990; Groden et al. 1991; Joslyn et al. 1991; Viskochil et al. 1990), we recognize that this strategy is both labor intensive and only moderately efficient. The inherent inadequacies of the strategy have been further accentuated by the efforts of various genome centers to physically map and sequence the entire genome and to position expressed sequences along each chromosome, which have resulted in a very extensive overlap with our efforts. This situation has been difficult for us to continue and has required us to redefine the means and methods necessary to achieve the original objective of our project; *to identify and characterize genes involved in breast cancer development and progression.* Having evaluated the situation, we determined that we would have to choose one of the following two experimental paths: a) determining the full-length sequence of previously mapped genes and screening them for mutations in breast tumors, or b) implementing a novel cancer gene identification strategy with a wide scope and of high efficiency. We chose to follow the latter path and in the following sections we describe the two approaches we have implemented to identify differentially expressed genes and to determine genetic expression profiles.

Statement of Work - revised July 28, 1997

Task 1, To apply the Microarray Scanning technique to identify genes displaying altered expression levels among breast cell lines.

During the past 3 years most of our efforts have been focused on the completion of Task 1 in our newly revised statement of work (SOW). As part of this effort is has been our primary goal to establish microarray analysis as a reliable and reproducible technology to compare gene expression profiles and identify differentially regulated genes. Work has been carried out in three areas being a) development of fluorescent labeling and hybridization protocols, b) selection and establishment of a collection of target genes, and c) gene expression comparisons between several different breast cell lines.

Development of probe labeling and hybridization protocols.

One of the most critical factors for a successful microarray experiment is the preparation of fluorescently labeled first strand cDNA to probe the microarray. The Molecular Dynamics Microarray System allows for dual color hybridization and the two fluorescent dyes we used are Cy3 and Cy5 from AP-Biotech. The labeling procedure itself is a multi step procedure that involves extraction of total RNA, mRNA purification and first strand cDNA synthesis. The fluorescent dyes conjugated to dCTP are incorporated into the cDNA to generate the fluorescent hybridization probe. We have found that SuperScript II results in better probes than AMV and MMLV reverse transcriptase, probably because it lacks proof reading ability. A variety of hybridization formats have also been tested. The most critical features of a successful hybridization with a complex probe

are to achieve and maintain high probe concentration during hybridization. To increase the probe concentration we have determined that the smallest practical hybridization volume is 20μl under a 22mm×60mm coverslip. Steps must also be taken to maintain proper salt concentrations in the hybridization buffer during incubation (i.e. eliminate evaporative effects). We have found that sealing the coverslip to the slide during hybridization leaves a fluorescent signature that affects the hybridization signal. In the procedure we currently employ the coverslip is not sealed, but the incubation takes place in a closed humidified chamber.

Selection and preparation of target cDNA clones.

A key reagent in microarray analysis is the collection of genes deposited on the array slide against which the levels of expression are measured. In our last annual report, we mentioned that we had obtained 23,000 minimally redundant cDNA clones from Genome Systems (St. Louis) and Research Genetics (Huntsville, Al). The procedures we have implemented to process this large number of clones are divided into multiple steps. In the first, mini cultures of each clone were grown and each plasmid DNA was purified in the 96-well plate format. Using the plasmid as a template in a PCR reaction in conjunction with vector based primers, the insert cDNAs were amplified. Following gel analysis and colorimetric analysis of the PCR yield, the PCR products were purified, still using the 96-well format, and deposited to the array slides. Although these 23,000 cDNA clones had been successfully used for our microarray experiments (such as Karpf et al. 1999, see attached reprint), we were later informed that about 30% of cDNA clones had been mislabeled by the company. However, fortunately during the past 6 months we were able to gain a new set of over 40,000 minimally redundant sequence-verified human cDNA clones from Research Genetics. These sequence-verified cDNA clones are now successfully processed and ready for the large-scale microarray experiments.

Analysis of primary breast cells with conditionally induced dominant β-catenin.

β-catenin has resently been shown to act as an oncogene in certain biologic models and to study the gene regulatory effects of the cancer pathway driven by Wnt signaling in breast cells, human primary epithelial mammary cells were transfected with a conditionally suppressible dominant-negative β-catenin construct as mentioned above. Excessive stimulation of this pathway and certain types of mutations have been shown to involve the formation of a persistent transcriptionally active complex of β-catenin and LEF1. mRNA from several different primary breast cultures, with and without induced β-catenin, was prepared and fluorescently labeled for dual color hybridization on microarray slides representative of 2400 unique cDNAs. Twenty-nine different genes have been identified in repeated experiments as differentially expressed and are listed in Table 1. To confirm that the microarray observations are correct Northern analysis of each differentially expressed gene is being undertaken. In Figure 1 we show that Psoriasin, which displayed a 3-4 fold increase in abundance judging from the microarray analysis reveal similar differences in abundance when tested by Northern analysis.

Statement of Work - revised July 28, 1997

Task 2, Characterization of growth and morphologic effects of conditionally expressed candidate cancer genes in cultured breast cells.

Comparison of three conditional expression vector systems.

To identify a conditional expression system we compared a dominant-negative β-catenin construct (N-terminal deletions, dN131 and dN151) in three different conditionally inducible/suppressible vectors, pRetro ON, pRetro OFF (Clontech) and LINX (Hoshimaru et al.

1996). These vector DNAs were transfected into the Phoenix Amphotropic packaging cell line (Dr. Garry Nolan) and the resulting infectious viral particles produced were used to infect primary human mammary epithelial cells. After two weeks of drug selection in either 0.5 μg/ml Puromycin (for pRetro ON and pRetro OFF) or 100 μg/ml G418 (for LINX), cells were treated with or without 10ng/ml doxycycline (DOX) for 48 hours. Total cell extracts were subjected to Western analysis using anti-C-terminal β-catenin antibody (Figure 2a). To demonstrate that induced dominant β-catenin is functional, cells were infected with LINX vector control and LINX with inducible dominant β-catenin gene (dN131) and subsequently transiently transfected with luciferase reporter construct, TOP FLASH, containing LEF1 responsive element in its promoter (obtained from Dr. Hans Clevers). As shown in Figure 2b, the relative activity of the LEF1-reporter construct showed 2.5 fold increase in activity following induction of β-catenin. Based on these results we have selected the LINX for future experimentation.

Characterization of growth and morphologic effects in cultured breast epithelial cells.

The consequences of mutations in regulatory genes, such as p53, Rb and BRCA1 genes, should be associated with observable cellular and molecular changes. Bisssell et al. have shown that culturing human mammary epithelial cells in an extracellular matrix called Matrigel provides a functionally relevant microenvironment conductive to form three-dimensional structures collagen (Weaver et al. 1996, Weaver et al. 1997). Normal cells will form differentiated spheroids with a central lumen. These structures express certain cell lineage markers such as sialomucin at the apical membrane and type IV collagen at the basal membrane. Breast cancer cells, however, are disorganized and lack polarized expression of these markers (Weaver et al. 1996, Weaver et al. 1997). In collaboration with the Bissell laboratory, we have established three-dimensional culture system in our laboratory. As an initial attempt to test the sensitivity of this culture system on morphology and growth due to the altered expression of tumor suppressor or growth control genes, we created human mammary epithelial cells (HMEC) deficient for pRB by infecting primary outgrowth from breast organoids with the human papillomavirus type 16 (HPV16) E7 gene (Spancake et al. 1999, see attached reprint). HPV16 E7 binds to and inactivates pRB, and also causes a significant down-regulation of the protein. Culturing normal HMEC in a reconstituted basement membrane (rBM) provides a correct environment and signaling cues for the formation of differentiated, acini-like structures. When cultured in this rBM, HMEC+E7 were found to respond morphologically as normal HMEC and form acinar structures. In contrast to normal HMEC, many of the cells within the HMEC+E7 structures were not growth arrested as determined by a BrdU incorporation assay. pRB deficiency did not affect polarization of these structures as indicated by the normal localization of the cell-cell adhesion marker E-cadherin and the basal deposition of a collagen IV membrane. However, in HMEC+E7 acini we were unable to detect by immunofluorescence microscopy the milk protein lactoferrin or cytokeratin 19, both markers of differentiation expressed in the normal HMEC structures. These data indicate loss of RB in vivo would compromise differentiation, predisposing these cells to future tumor-promoting actions, suggesting that this culture system would be suited to examine the function of other tumor suppressor and growth control genes.

(7) KEY RESEARCH ACCOMPLISHMENTS

- DLG2, DLG3, and ARF4L genes were mapped to BRCA1 flanking region. However , we have no evidence so far to indicate that either DLG2, DLG3 or ARF4L is the primary target in the region of LOH flanking BRCA1.
- We have established the microarray spotting and scanning system with over 40,000 minimally redundant sequence-verified human cDNA clones.
- We have established a reliable retroviral-based conditional expression system.
- We have established a system to examine the growth and morphological properties of human breast epithelial cells cultured in an extracellular matrix (Matrigel).

(8) REPORTABLE OUTCOMES

Manuscripts

Mazoyer S, Gayther SA, Nagai MA, Smith SA, Dunning A, van Rensburg EJ, Albertsen H, White R, & Ponder BA. A gene (DLG2) located at 17q12-q21 encodes a new homologue of the Drosophila tumor suppressor dIg-A. Genomics. 1995 Jul 1;28(1):25-31.

Smith SA, Holik PR, Stevens J, Melis R, White R, & Albertsen H. Isolation and mapping of a gene encoding a novel human ADP-ribosylation factor on chromosome 17q12-q21. Genomics. 1995 Jul 1;28(1):113-5.

Brothman AR, Steele MR, Williams BJ, Jones E, Odelberg S, Albertsen HM, Jorde LB, Rohr LR, & Stephenson RA. Loss of chromosome 17 loci in prostate cancer detected by polymerase chain reaction quantitation of allelic markers. Genes Chromosomes Cancer. 1995 Aug;13(4):278-84.

Smith SA, Holik P, Stevens J, Mazoyer S, Melis R, Williams B, White R, & Albertsen H. Isolation of a gene (DLG3) encoding a second member of the discs-large family on chromosome 17q12-q21. Genomics. 1996 Jan 15;31(2):145-50.

Albertsen HM, Smith SA, Melis R, Williams B, Holik P, Stevens J, & White R. Sequence, genomic structure, and chromosomal assignment of human DOC-2. Genomics. 1996 Apr 15;33(2):207-13.

Karpf AR, Peterson PW, Rawlins JT, Dalley BK, Yang Q, Albertsen H, Jones DA. Inhibition of DNA methyltransferase stimulates the expression of signal transducer and activator of transcription 1, 2, and 3 genes in colon tumor cells. Proc Natl Acad Sci U S A. 1999 Nov 23;96(24):14007-12.

Spancake KM, Anderson CB, Weaver VM, Matsunami N, Bissell MJ, & White RL. E7-transduced human breast epithelial cells show partial differentiation in three-dimensional culture. Cancer Res. 1999 Dec 15;59(24):6042-5.

(9) CONCLUSIONS

During the funded period, the overall goal of our project is to identify genes involved with the development and progression of breast cancer. This goal has remained unchanged since the start of the project, but the discovery of BRCA1 in 1994 together with technological advances in gene expression profiling has influenced our strategy to achieve this goal. In the early part of the project our search for tumor suppressor genes was directed by genetic or LOH mapping strategies followed by positional cloning of candidate genes. As we proposed in our revised statement of work (SOW), we have focused our efforts entirely on the development of microarray-based comparisons to identify breast cancer related genes and further characterize the change in morphology and growth

of human mammary epithelial cells conditionally expressing the breast cancer related genes using the Matrigel culture system.

(10) REFERENCES

H. Aberle, C. Bierkamp, D. Torchard, O. Serova, T. Wagner, E. Natt, J. Wirsching, C. Heidkamper, M. Montagna, H.T. Lynch, G.M. Lenoir, G. Scherer, J. Feunteun and R. Kemler. "The human plakoglobin gene localizes to chromosome 17q21 and is subject to loss of heterozygosity in breast and ovarian cancers". *Proc. Natl. Acad. Sci. USA* **92**: 6384-6388.

H.M. Albertsen et al. "Genetic mapping of the BRCA1 region on chromosome 17q21." *Am. J. Hum. Genet.* **54**:516-525 (1994a).

H.M. Albertsen, S.A. Smith, S. Mazoyer, E. Fujimoto, J. Stevens, B. Williams, P. Rodriguez, C.S. Cropp, P. Slijepcevic, M. Carlson, M. Robertson, P. Bradley, E. Lawrence, Z.M.Sheng, R. Hoopes, N. Sternberg, A. Brothman, R. Callahan, B.A.J. Ponder and R. White. "A physical map and candidate genes in the *BRCA1* region". *Nature Genetics* **7**: 472-479 (1994b).

H.M Albertsen, S.A. Smith, R. Melis, B. Williams, P. Holik, J. Stevens, and R. White. "Sequence, genomic structure, and chromosomal assignment of human DOC-2." *Genomics* **33**:207-213 (1996).

A.R. Brothman, M.R. Steele, B.J. Williams, E. Jones, S. Odelberg, H.M. Albertsen, L.B. Jorde, L.R. Rohr and R. Stephenson. "Loss of chromosome 17 loci in prostate cancer detected by polymerase chain reaction quantitation of allelic markers". *Genes Chromosomes and Cancer* **13**: 278-284 (1995).

M.A. Brown, K.A. Jones, H. Nicolai, M. Bonjardim, D. Black, R. McFarlane, P. de Jong, J. P. Quirk, H. Lehrach and E. Solomon. "Physical mapping, cloning, and identification of genes within a 500-kb region containing BRCA1". *Proc. Natl. Acad. Sci. USA* **92**: 4362-4366 (1995).

L.H. Castilla, F.J. Couch, M.R. Erdos, K.F. Hoskins, K. Calzone, J.E. Garber, J. Boyd, M.B. Lubin, M.L. Deshano, L.C. Brody, et al "Mutations in the BRCA1 gene in families with early-onset breast and ovarian cancer." *Nature Genetics* **8**: 387-91 (1994).

R. Cawthon R et al. A major segment of the neurofibromatosis type 1 gene: cDNA sequence, genomic structure, and point mutations. Cell **62**:193-201 (1990).

I.M. Chumakov and 61 co-authors. "A YAC contig map of the human genome". *Nature* **377**: 175-297 (1955) (Supplement)

M. Coss, D. Winterstein, Rn. Sowder, S. Simek. Molecular cloning, DNA sequence analysis, and biochemical characterization of a novel 65-kDa FK506-binding protein (FKBP65). J Biol Chem **270**:29336-41 (1995).

C.S. Cropp, Z.-M. Sheng, H.A. Nevanlinna, H. Albertsen, R. White and R. Callahan "Genomic location and structure of plakoglobin". *-in preparation.*

C.S. Cropp, H.A. Nevanlinna, S. Pyrhönen, U.-H. Stenman, P. Salmikangas, H. Albertsen, R. White and R. Callahan. "Evidence for involvement of BRCA1 in sporadic breast carcinomas" *Cancer Research* **54**: 2548-2551 (1994).

L.S. Friedman, E.A. Ostermeyer, C.I. Szabo, P. Dowd, E.D. Lynch, S.E. Rowell and M.C. King. "Confirmation of BRCA1 by analysis of germline mutations linked to breast and ovarian cancer in ten families". *Nature Genetics* **8**: 399-404 (1994).

P.A. Futreal, Q. Liu, D. Shattuck-Eidens, C. Cochran, K. Harshman et al. "BRCA1 mutations in primary breast and ovarian carcinomas". *Science* **266**: 120-122 (1994).

S.C. Gerken, H. Albertsen, T. Elsner, L. Ballard, P. Holik, E. Lawrence, M. Moore, X. Zhao and R. White. "A strategy for constructing high-resolution genetic maps of the human genome: A genetic map of choromosome 17p ordered with meiotic breakpoint mapping panels." *American Journal of Human Genetics* **56**: 484-499 (1995).

J. Groden et al. Identification and characterization of the familial adenomatous polyposis coli gene. Cell **66**:589-600 (1991).

M. Hoshimaru, J. Ray, D. Sah, F. Gage. Differentiation of the immortalized adult neuronal progenitor cell line HC2S2 into neurons by regulatable suppression of the v-myc oncogene. *Proc. Natl. Acad. Sci. USA* **93**:1518-23 (1996).

G. Joslyn et al. Identification of deletion mutations and three new genes at the familial polyposis locus. Cell **66**:601-13 (1991).

A.R. Karpf, P.W. Peterson, J.T. Rawlins, B.K. Dalley, Q. Yang, H. Albertsen, and D.A. Jones. Inhibition of DNA methyltransferase stimulates the expression of signal transducer and activator of transcription 1, 2, and 3 genes in colon tumor cells. *Proc. Natl. Acad. Sci. USA* **96**:14007-14012 (1999).

C. Kirkpatrick and M. Peifer. " Not just glue: cell-cell junctions as cellular signaling centers". *Current Opinion in Genetics and Development* **5**: 56-65 (1995).

S.M. Mazoyer, S.A. Gayther, M.A. Nagai, S.A. Smith, A. Dunning, E.J. van Rensburg, H. Albertsen, R. White and B.A.J. Ponder. "A gene located at 17q12-21 encodes a new homologue of the Drosophila tumor-suppressor dlg-A protein" *Genomics* **28**: 25-31 (1995).

Y. Miki, J. Swensen, D. Shattuck-Eidens, P.A. Futreal, K. Harshman, S. Tavtigian, Q. Liu, C. Cochran, L.M. Bennett, W. Ding, et al. "A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1". *Science* **266**: 66-71 (1994).

J.C. Murray, K.H. Buetow, J.L. Weber, S. Ludwigsen, T. Scherpbier-Heddema, F. Manion, J. Quillen, V.C. Sheffield, S. Sunden, G.M. Duyk, J. Weissenbach, G Gyapay, C. Dib, J. Morrissette, G.M. Lathrop, A. Vignal, R. White, N. Matsunami, S. Gerken, R. Melis, H. Albertsen, K. Ward, R. Plaetke, S. Odelberg, D. Ward, J. Dausset, D. Cohen, H. Cann. "A comprehensive human linkage map with centimorgan density". *Science* **265**: 2049-2054 (1994).

J. Simard, P. Tonin, F. Durocher, K. Morgan, J. Rommens, S. Gingras, C. Samson, J.F. Leblanc, C. Belanger, F. Dion et-al "Common origins of BRCA1 mutations in Canadian breast and ovarian cancer families". *Nature Genetics* **8**: 392-398 (1994).

S.A. Smith, P.R. Holik, J. Stevens, R. Melis, R. White and H. Albertsen. "Isolation and mapping of a gene encoding a novel human ADP-ribosylation factor on chromosome 17q12-q21." *Genomics* **28**: 113-115 (1995).

S.A. Smith, P.R. Holik, J. Stevens, S. Mazoyer, R. Melis, B. Williams, R. White and H. Albertsen. "Isolation of a gene (DLG3) encoding a second member of the disc-large family on chromosome 17q12-21." *Genomics* **31**:145-150 (1996).

C.L. Sommers, E.P. Gelman, R. Kemler, P.Cowin, and S.W. Byers "Aletrations in -Catenin Phosphorylation and Plakoglobin Expression in Human Breast Cancer Cells". *Cancer Res.* **54**: 3544-3552 (1994).

K.M. Spancake, C.B. Anderson, V.M. Weaver, N. Matsunami, M.J.Bissell, and R.L.White. E7-transduced human breast epithelial cells show partial differentiation in three-dimensional culture. *Cancer Res.* **59**:6042-6045 (1999).

L.K. Su, B. Vogelstein, K.W. Kinzler. "The APC tumor suppressor protein associates with catenins". *Science* **262**: 1734-1737 (1993).

B.J. Williams, E. Jones, X.J. Zhu, M.R. Steele, R.A. Stephenson, L.R. Rohr and A.R. Brothman. "Evidence for a tumor suppressor gene distal to BRCA1 prostate cancer." *Urology* (submitted) (1995).

THE UTAH MARKER DEVELOPMENT GROUP: D. Adamson, H. Albertsen, L. Ballard, P. Bradley, M. Carlson, P. Cartwright, T. Elsner, D. Fuhrman, S. Gerken, L. Harris, P. R Holik, A. Kimball, J. Knell, E. Lawrence, J. Lu, A. Marks, N. Matsunami, R. Melis, B. Milner, M. Moore, L. Nelson, S. Odelberg, G. Peters, R. Plaetke, R. Riley, M. Robertson, R. Sargent, G. Staker, A. Tingey, K. Ward, X. Zhao and R. White. "A collection of ordered tetra-nucleotide repeat markers from the human genome." *American Journal of Human Genetics* **57**: 619-628 (1995).

D.F. Woods and P.J. Bryant. "Molecular cloning of the lethal (1) discs-large (1) oncogene of Drosophila". *Dev. Biol.* **134**: 222-235.

V. Weaver, A, Fischer, O. Peterson, and M. Bissell. The importance of the microenvironment in breast cancer progression: recapitulation of mammary tumorigenesis using a unique human mammary epithelial cell model and a three-dimensional culture assay. Biochem Cell Biol **74**:833-51 (1996).

V, Weaver, O. Petersen, F, Wang, C. Larabell, P. Briand, C. Damsky, and M. Bissell. Reversion of the malignant phenotype of human breast cells in three-dimensional culture and in vivo by integrin blocking antibodies. J Cell Biol **137**:231-45 (1997).

D. Viskochil et al. Deletions and a translocation interrupt a cloned gene at the neurofibromatosis type 1 locus. Cell **62**:187-92 (1990).

(11) APPENDICES

Materials and methods

Cell Culture and retroviral infection.

Surgical discard material from reduction mammoplasties was minced with opposing scalpels, placed in digestion buffer, and incubated in spinner flasks at 37°C until stroma dissolved (approximately three to five hours). Digestion buffer contained 1 unit/ml Collagenase D (Roche Molecular Biochemicals, Indianapolis, IN), 2.4 units/ml dispase (Roche Molecular Biochemicals), and 6.25 units/ml DNase (Sigma, St. Louis, MO) in Dulbecco's phosphate buffered saline. The digested material plus 10% fetal calf serum (FCS) was centrifuged at 800 rpm for 10 min. The resulting pellet was resuspended in wash buffer and the organoids were separated from stromal and blood cells by sequential sedimentation at 1x g. Organoids were cultured immediately or frozen in Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (Ham) 1:1 (GibcoBRL, Gaithersburg, MD), 10% FCS, and 10% dimethyl sulphoxide. Organoids were cultured on plastic in CDM3 culture media for primary breast epithelial cell outgrowth and subculture. Primary epithelial outgrowth was infected with an LXSN retroviral construct containing the human papilloma virus type 16 E7 gene (LXSN16E7). HMEC were incubated with LXSN16E7 in CDM3 plus 4 μg/ml polybrene (Sigma) for 24 h. Viral supernatant was aspirated and HMEC were cultured in virus-free CDM3 for 48 h prior to selection in CDM3 containing 50 μg/ml Geneticin (Gibco BRL). Early passage HMEC (either passage one or two) and HMEC containing LXSN16E7 (HMEC+E7) were cultured in a rBM, Matrigel (Becton Dickinson, Bedford, MA), as previously described. Briefly, 2.5 x 10^5 HMEC were resuspended as single cells in 300 μl of 10 mg/ml Matrigel per well and plated onto Nunc 4-well multidishes coated with 100 μl Matrigel. Matrigel cultures were overlaid with 500 μl CDM3.

Table 1.

	Image ID	Clone definition
1	358433	Human retinoid X receptor-gamma mRNA, complete cds
2	295401	ESTs
3	1088345	S100 calcium-binding protein A7 (psoriasin 1)
4	276282	ESTs
5	269017	Human O-linked GlcNAc transferase mRNA, complete cds
6	545239	Neutrophil Gelatinase-Associated Lipocalin Precursor
7	882141	DNA G/T mismatch-binding protein
8	283063	MHC class II DQ-beta associated with DR2, DQw1 protein
9	156431	Ciliary neurotrophic factor receptor
10	770435	Transcription factor p65
11	700466	
12	277134	ESTs
13	275272	ESTs
14	273039	ESTs
15	283618	ESTs
16	23240	SM22-alpha homolog
17	627104	Human alpha-tubulin mRNA, complete cds
18	33934	ESTs
19	610187	Proteasome Component C9
20	592947	Autoantigen PM-SCL
21	46743	ESTs, Highly similar to RAS-related protein RAP-1B
22	23904	ESTs
23	264369	ESTs, Highly similar to GLUCOSYLTRANSFERASE ALG8
24	200531	ESTs
25	126783	ESTs
26	201891	ESTs
27	265684	ESTs
28	261971	Metallopeptidase 1 (33 kD)
29	129503	ESTs

The genes listed in this table showed 3× fold or more variation in expression levels in a comparison between primary human mammary cells, with and without induced β-catenin.

Figure 1.
This figure shows a Northern Blot of four different cell lines transfected with the LINX β-catenin construct and the vector control. In absence of Doxycilin (indicated by -) β-catenin is over-expressed resulting in a three fold up-regulation Psoriasin.

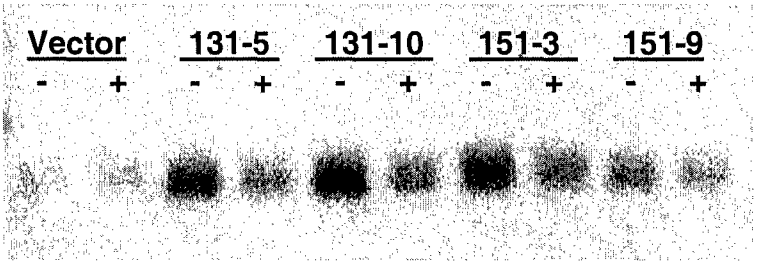
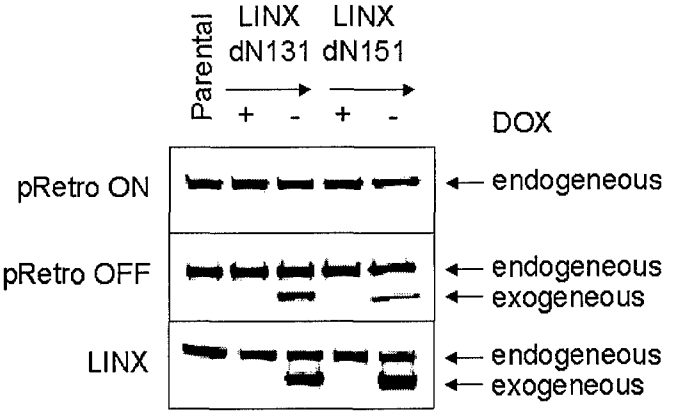


Figure 2.

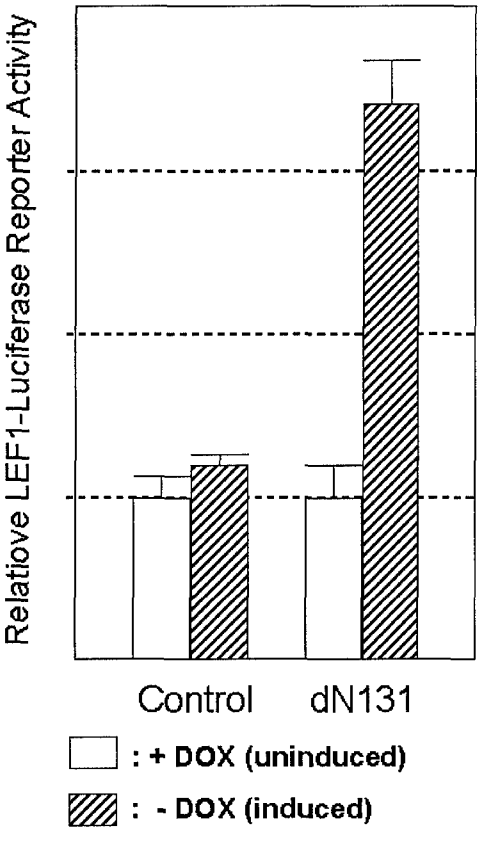
a. Conditionally inducible vectors.

To test the efficacy of the tetracycline regulated retroviral vector system, dominant-negative b-catenin genes were cloned into pRetro Off, pRetro On, and LINX. Following packaging, transfection into primary human mammary epithelial cells and two weeks of selection, total cell extracts were subjected to Western analysis using anti-C-terminal β-catenin antibody. As shown here, cells infected with pRetro ON construct did not induce dominant β-catenin at all. Cells infected with pRetro OFF induced dominant β-catenin in response to DOX removal, however the amount induced (shown as exogeneous) were much less than that of endogeneous protein. On the other hand, cells infected with LINX vector construct produced great response in both inducibility and amount.



b. LEF1-Luciferase Reporter Assay.

Shown here are relative LEF1-reporter activities indicating that induced dominant β-catenin is in fact functional in regard to activating genes containing LEF1 responsive element.



GENES, CHROMOSOMES & CANCER 13:278–284 (1995)

Loss of Chromosome 17 Loci in Prostate Cancer Detected by Polymerase Chain Reaction Quantitation of Allelic Markers

Arthur R. Brothman, Michael R. Steele, Briana J. Williams, Emma Jones, Shannon Odelberg, Hans M. Albertsen, Lynn B. Jorde, L. Ralph Rohr, and Robert A. Stephenson

Departments of Human Genetics (A.R.B., S.O., H.M.A., L.B.J.), Pediatrics (A.R.B., M.R.S., B.J.W., E.J.), Pathology (L.R.R.), and Urology (R.A.S.), University of Utah School of Medicine, Salt Lake City, Utah

Using a polymerase chain reaction/microsatellite marker system, we demonstrated that 6 of 22 (27%) clinical stage B (early) primary prostate tumors showed loss of heterozygosity at one or more of five loci on chromosome 17. The sensitivity of this study was increased by use of a PhosphorImager and statistical analysis of replicate tumor-normal DNA pairs. Two patients showed tumor-specific interstitial loss at a locus in close proximity to the familial breast cancer gene *BRCA1*. These findings suggest that genes on the proximal long arm of chromosome 17 play a pivotal role in the early development of at least a subset of prostatic tumors. *Genes Chromosom Cancer 13:278–284 (1995).*

INTRODUCTION

Prostate cancer is the most common cancer in males in the United States, with an estimated 200,000 new cases in 1994 (Boring et al., 1994), yet the etiology of this disease is still poorly understood. Genetic and cytogenetic changes have been associated with the disease, but no consistent cellular abnormality has been observed. The recent use of fluorescence in situ hybridization (FISH) techniques with chromosome-specific probes has suggested that chromosome 17 (i.e., the whole chromosome) is frequently lost in prostate tumors (Brothman et al., 1992, 1994; Jones et al., 1994); hence, we sought to evaluate primary tumor tissue for molecular abnormalities of this chromosome.

Several molecular studies have been published in which certain chromosomal regions are lost in prostatic tumors, suggesting that loss of a tumor suppressor function may be involved in this disease. Loss of loci on the short arm of chromosome 8 is, to date, the most frequently observed loss in prostate tumors (Bergerheim et al., 1991; Kunimi et al., 1991; Bova et al., 1993; Macoska et al., 1993). A putative tumor suppressor gene in this region has been associated with other solid tumors, including hepatocellular, colorectal, and lung cancers (Emi et al., 1992, 1993). Chromosomal loss has also been observed in prostate cancer at sites including 5q, 10p, 10q, 16q, 17p, and 18q (Carter et al., 1990; Bergerheim et al., 1991; Kunimi et al., 1991; Brewster et al., 1994; Latil et al., 1994). Macoska and colleagues (1992) have shown that there is loss of short-arm material from chromosome 17 in metastatic prostate tumors, and mutations in the *TP53* gene have also been observed in a rare subset of advanced prostate tumors (Effert et al., 1992; Bookstein et al., 1993). Recently, Brewster and colleagues (1994) have observed loss at the nm23-H1 locus on 17q in one late-stage prostate tumor; however, no reports have indicated loss of regions on the long arm of chromosome 17 in early-stage tumors.

The susceptibility gene for familial breast and ovarian cancer, *BRCA1*, is located on the proximal long arm of chromosome 17 (Miki et al., 1994) and has been theorized to be involved in prostate cancer (Arason et al. 1993; Ford et al., 1994). Therefore, we chose to study markers near *BRCA1* in our initial screen for loss of heterozygosity (LOH) on chromosome 17, to our knowledge, previous studies included only markers on distal 17q.

The considerable cellular heterogeneity within the prostate gland creates difficulties in allelotyping of these tumors. Some investigators have used microdissection of apparent tumor tissue for DNA preparation (Bova et al., 1993; Macoska et al., 1993; MacGrogan et al., 1994), whereas others have limited their analysis to gross pathologic characterization prior to DNA extraction (Bergerheim et al., 1991). Even with the most careful microdissections, however, some proportion of contaminating normal epithelial cells or of fibroblastic stromal cells remains. We have approached the problem of

Received November 29, 1994; accepted March 31, 1995.

Address reprint requests to Arthur R. Brothman, PhD, Departments of Pediatrics and Human Genetics, 1C204 University of Utah Medical Center, University of Utah School of Medicine, Salt Lake City, UT 84132, U.S.A.

cellular heterogeneity in our specimens by relying on the gross histopathologic evaluation of adjacent tumor specimens and by analyzing multiple replications of polymerase chain reaction (PCR) products with a Molecular Dynamics PhosphorImager. Using this technique, we have observed tumor-specific losses of regions on chromosome 17 in 6 of 22 (27%) of the patients evaluated. Furthermore, we have confirmed these results by using a novel FISH approach for deletion detection with selected single-copy P1 clones serving as probes (Williams et al., 1995).

MATERIALS AND METHODS

Specimens

Twenty-two primary prostate tumor specimens obtained from radical prostatectomies at the University of Utah School of Medicine were prepared for DNA and single-cell preparations. A central section of a region of nodularity was obtained at surgery, and an adjacent section was evaluated histopathologically as described previously (Jones et al., 1994). All specimens were from patients who had early (clinical stage B) cancer at evaluation. Tumor specimens were coded and separated for various other ongoing studies in this laboratory, and portions were frozen in liquid nitrogen for subsequent DNA extraction. A peripheral blood sample was obtained prior to surgery from each patient for direct preparation of constitutional DNA and for immortalization to a lymphoblastoid line.

DNA Extraction

Frozen tissue was ground with a micropestle, incubated in a lysis buffer (10 mM Tris, pH 8.0, 100 mM NaCl, 1 mM EDTA, 1% SDS, and 0.2 mg/ml proteinase K) at 37°C overnight, phenol-chloroform/isoamyl alcohol extracted, and ethanol precipitated prior to resuspension in 10 mM Tris/0.1 mM EDTA buffer. DNA was extracted from peripheral blood that was first treated with Triton X lysis buffer (0.32 M sucrose, 10 mM Tris-HCl, pH 7.5, 5 mM $MgCl_2$, 1% Triton X-100), followed by treatment of pelleted nuclei with a buffer containing 1% SDS/0.075 M NaCl/0.024 M EDTA, pH 8.0, and 0.25 mg/ml proteinase K overnight at 37°C. DNA was isolated the following day by phenol and chloroform/isoamyl alcohol extraction and ethanol precipitation prior to resuspension in Tris-EDTA buffer (Bell et al., 1981).

Microsatellite Primer Sets

Primer sets for the following markers were generated through the Utah Genome Center Marker Development and Mapping Group: UT7 (D17S752), UT573 (D17S902), and UT752 (D17S907) located on the long arm of chromosome 17 (Utah et al., in press), and UT751 (D17S906) and UT5265 (D17S1149) located on the short arm of chromosome 17 (Albertsen et al., 1994). In addition, a polymorphic marker for the short arm of chromosome 8 at the *LPL* locus (Zuliani and Hobbs, 1990) was analyzed for each tumor-normal DNA pair for information on an independent chromosome and as a test for efficiency of our PCR technique in comparison to frequencies of loss observed at this locus by others.

PCR

One member of a specific primer set was labeled at the 5′ end with [^{32}P]gamma ATP (New England Nuclear-Dupont) by use of polynucleotide kinase (Boehringer Mannheim). The PCR was optimized and performed with a Techne PHC-3 thermal cycler with a mixture of each primer set (1:100 radioactively labeled to unlabeled primer), dNTPs (Pharmacia), 0.001 U/μl Taq polymerase (Boehringer Mannheim), 2 mM spermidine, and PCR buffer (10 mM Tris-HCl, pH 8.8, 40 mM NaCl containing specific concentrations of Mg^{2+} ranging from 1 to 2 mM).

Analysis of PCR Product

After amplification of the PCR product, the samples were run on an acrylamide denaturing gel containing 7% acrylamide [38:1 acrylamide:bisacrylamide (BioRad), 32% formamide (BRL), 5.6 M urea, and Tris-borate-EDTA buffer] and exposed overnight at −80°C in an X-ray film holder with intensifying screen and Amersham Hyperfilm MP. Images were both evaluated visually and analyzed with a Molecular Dynamics PhosphorImager. After exposure to X-ray film, the gel was placed against a PhosphorImage screen for 20 minutes to 16 hour, depending on signal intensity. A digital image was obtained and was then analyzed by the drawing of boundaries around each allelic band, subtraction of local background, and integration of the volume of each band. The value of the top band was always taken as a percentage of the bottom band.

Statistical Analysis

At least four replicates (separate aliquots of the same DNA isolates) of each tumor and the corresponding normal sample were run in parallel on the PhosphorImager, and differences were then evaluated for significance by use of the nonparametric Mann-Whitney U test. This is considered a robust

TABLE 1. Control Mixing Experiment[a]

	95	90	85	80	75	70	65	60	55	50	45	40	35	30	25	20	15	10	5
100	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
95		−	−	−	−	−	+	+	+	+	+	+	+	+	+	+	+	+	+
90			−	−	−	−	+	+	+	+	+	+	+	+	+	+	+	+	+
85				−	−	+	+	+	+	+	+	+	+	+	+	+	+	+	+
80					−	−	+	+	+	+	+	+	+	+	+	+	+	+	+
75						−	−	−	−	−	+	+	+	+	+	+	+	+	+
70							+	+	−	+	+	+	+	+	+	+	+	+	+
65								−	−	−	+	+	+	+	+	+	+	+	+
60									−	−	+	+	+	+	+	+	+	+	+
55										−	+	+	+	+	+	+	+	+	+
50											−	+	+	+	+	+	+	+	+
45												+	+	+	+	+	+	+	+
40													−	+	+	+	+	+	+
35														+	−	+	+	+	+
30															−	+	+	+	+
25																+	+	+	+
20																	+	+	+
15																		−	+
10																			+

[a]Comparison of percent-mixed groups of normal (two-allele) DNA to increasing percentages of the tumor (one-allele) DNAs. Values represent percent *normal* DNA; + and − signify difference or no difference, respectively, by the Mann-Whitney U test for eight replicates when PhosphorImager data at the percentages shown on the x and y coordinates of this table were compared.

test and is preferable to a parametric Student's t test, because it does not assume a normal distribution. Four replicates were used for all initial evaluations; if overlapping values were detected for allele ratios between tumor and normal specimens, then additional replicates were evaluated applying the Mann-Whitney U test to all accumulated values. We performed a power analysis to determine the probability of incorrectly rejecting the null hypothesis of no difference (presence of LOH, type I error) and the probability of incorrectly accepting the null hypothesis of no difference (no LOH, type II error). If the type I error is fixed at 0.05 (i.e., the significance level for deciding that an allele has been lost), then the Mann-Whitney U test yields type II error levels of 0.11 when eight replicates are used and 0.35 when four replicates are used.

Controlled experiment for LOH

To ascertain the sensitivity of the PhosphorImager quantitation, we performed a controlled mixing experiment by using DNA from a lymphoblastoid line from a patient who was informative for marker UT573 (analogous to "normal" DNA) and DNA from cell line MH-22.6 (Coriell Institute for Medical Research, Camden, NJ). This human/rodent somatic cell hybrid line contains a single copy of human chromosome 17 and, thus, shows only one allele when primers for UT573 are used (analogous to "tumor" DNA in the experiment). The patient whom we chose for lymphoblastoid DNA contained an allele at UT573 identical to that of cell line MH-22.6. These DNAs were combined to simulate normal DNA contamination of tumor DNA at 5% increments from 0 to 100%. UT573 primers were used for PCR amplification and subsequent PhosphorImager analysis as described above; eight replicates (from separate PCR reactions) were tested for each mixture.

RESULTS

Controlled Mixing Experiment

The results of the simulated "tumor-normal" cell contamination are shown in Table 1. Whereas we were able to detect allelic imbalance in this experiment with 95% contaminating normal DNA, the experiment in Table 1 was designed to replicate other unforeseen variables such as differential allelic intensities. This was done by extending our analyses to decreasing amounts of tumor DNA compared to *less than 100%* "control" or normal DNA. Although allelic imbalance was generally detected in high amounts of normal DNA, the lowest observed imbalance was seen at 60% contaminating normal (75% value, which did not detect allelic loss until at least 45% of normal DNA was present;

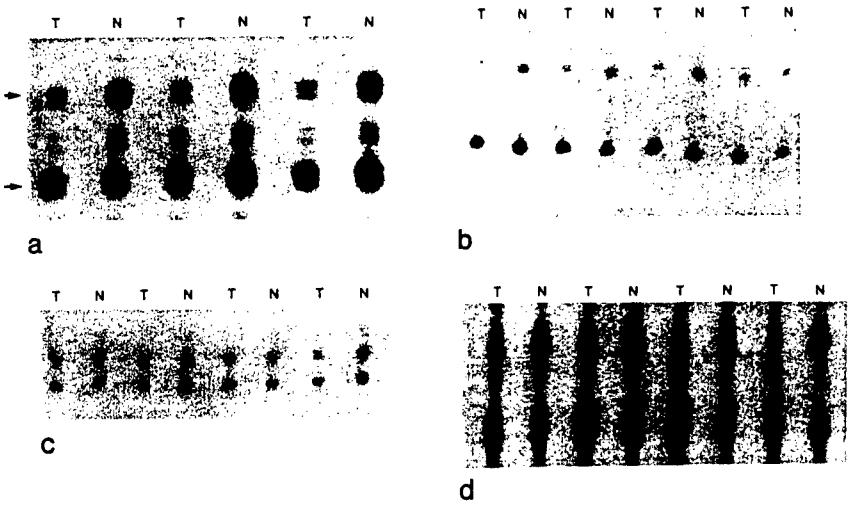


Figure 1. Representative autoradiographs of replicates showing LOH in PCR products at different loci. The various markers and corresponding specimens (UCAP number) are as follows. **a:** *LPL* (UCAP 14). Arrows indicate allelic bands. **b:** UT751 (UCAP 18). **c:** UT752 (UCAP 24). **d:** UT573 (UCAP 24).

TABLE 2. Summary of Informativeness and LOH Data for Markers Used

Marker	Location	Heterozygous (%)	LOH (%)
LPL	8p22	13 of 22 (59)	5 of 13 (39)
UT7	17q21.3	14 of 22 (64)	1 of 14 (7)
UT573	17q21.2	16 of 22 (73)	4 of 16 (25)
UT752	17q11.2-12	18 of 22 (82)	4 of 18 (22)
UT751	17p12-13	19 of 22 (86)	3 of 19 (16)
UT5265	17p12-13	15 of 22 (68)	4 of 15 (27)

45%/75% = 0.6). Overall, these results confirm our prediction that we can detect loss in a highly contaminated background. The relatively conservative Mann-Whitney U test criteria were used for the following analysis of prostate tumors.

LOH

Tumor-specific allelic losses on chromosome 17 were observed in 6 of 22 prostate cancer specimens (27%) with the markers used in this study. A summary of informativeness (frequency of heterozygosity) of the markers used and the LOH observed is shown in Table 2. Representative autoradiographs of replicates showing varying degrees of loss at particular loci for individual patients are presented in Figure 1, with corresponding PhosphorImager data in Table 3. The numerical data from the PhosphorImager analysis were chosen because they provide an objective measurement for allelic loss and an increased sensitivity over visual detection, which can be seen when data from Figure 1 and Table 3 are compared. When all data were evaluated, loss of the large (upper band) or small (lower band) allele was seen, which is expected in any allelic imbalance study. If there was any overlap in values where ratio differences were suspected (seen for UCAP 24 with the marker UT752 in Table 3), then we ran additional PCR amplifications and gels to confirm or refute our initial observations.

A summary of these data is shown in Table 4 and is depicted in the partial map of chromosome 17 in Figure 2. Values in Table 4 represent mean allelic ratios determined by PhosphorImager analysis. Hence, in general, the closer these values are to 100%, the more likely it is that tumor and normal values were identical. It is important to remember, however, that the statistic used (Mann-Whitney U test) also takes into account the possibility of overlap in values (see the example discussed above). Because we have chosen a significance level of $P \leq 0.05$, we are 95% confident that LOH detected represents true loss. Values that appear to deviate from 100% and were not significant (e.g., UCAP 5 at marker UT5265) represent mean values, but a wider range of PhosphorImager data deemed these loci insignificant.

It can be seen in Figure 2 and Table 4 that two patients showed interstitial loss on 17q (UCAP 24 and UCAP 30), one patient showed an apparent loss of the entire chromosome 17 (UCAP 18), one patient showed only short-arm loss (UCAP 17), and one patient showed loss at both distal ends of chromosome 17 (UCAP 29). Loss of UT573 was also seen in UCAP 28, whereas the most centromeric long-arm marker was retained. Because the distal marker, UT7, was uninformative for that specimen, we have not considered this to be an interstitial loss.

For a positive control, primers at an independent chromosome 8 locus were tested in a similar manner. Loss at the *LPL* locus on chromosome 8 was observed in five cases (39% of informative cases)

TABLE 3. Representative PhosphorImager Allele Ratios for Corresponding Photographs in Figure 1[a]

Locus (Specimen)	Allele ratio, tumor	Allele ratio, normal
LPL (UCAP 14)	27.9/25.4/32.2	67.8/70.1/71.4
UT751 (UCAP 18)	30.1/31.6/34.0/37.1	51.0/54.4/60.8/50.3
UT752 (UCAP 24)	91.2/105.8/106.3/82.4	103/114.2/104.9/112.8[b]
UT573 (UCAP 24)	47.3/51.2/54.2/53.8	92.9/101.9/100.6/96.3

[a]Corresponding numbers in each series represent adjacent lanes in Figure 1.
[b]Because of the overlapping values in this series, additional replicates were evaluated, and the statistical power was thus increased (final analyses for UCAP 24, see Table 4).

TABLE 4. Primary Prostate Tumor/Normal PhosphorImager Data and Significant Loss at Five Chromosome 17 (UT) Markers and One Chromosome 8 (LPL) Marker

	Microsatellite markers					
Specimen	UT7	UT573	UT752	UT751	UT5265	*LPL*
UCAP 2	N	N	96	84	N	99
UCAP 3	94	N	99	N	88	N
UCAP 4	89	N	93	N	99	N
UCAP 5	94	98	N	99	60	N
UCAP 6	97	93	92	93	66	96
UCAP 14	89	92	99	94	96	47[a]
UCAP 16	N	99	85	79	N	N
UCAP 17	N	94	N	43[a]	57[a]	57[a]
UCAP 18	N	87[a]	75[a]	61[a]	79[a]	N
UCAP 19	99	98	98	99	99	N
UCAP 20	91	95	93	97	86	N
UCAP 21	88	99	99	99	83	94
UCAP 23	97	98	88	67	N	99
UCAP 24	96	54[a]	82[a]	98	N	52[a]
UCAP 25	89	95	98	95	N	N
UCAP 26	N	92	93	96	93	94
UCAP 27	N	N	N	83	93	76
UCAP 28	N	54[a]	96	49[a]	42[a]	50[a]
UCAP 29	90[a]	N	96	98	69[a]	N
UCAP 30	95	83[a]	N	96	94	99
UCAP 31	100	94	92	99	N	95
UCAP 33	N	N	90	N	N	63[a]

[a]LOH. Numbers represent percentages of the normal vs. tumor allele ratios for the mean values for each informative set of replicates [(allelic tumor ratios/allelic normal ratios) × 100]. Significance values ($P \leq 0.05$) show that tumor allele ratios differ from normal allele ratios using the Mann-Whitney U test. In *general*, the closer a number is to 100%, the closer *all* replicate PhosphorImager values were between tumor and normal pairs. N, noninformative.

including UCAP 18, UCAP 24, and UCAP 28, which showed loss of loci on chromosome 17 as well.

DISCUSSION

The results of this study suggest that deletions or LOH on the long arm of chromosome 17 may be involved in some early-stage prostate tumors. Although allelic loss on 17q has not been reported previously in early-stage prostate tumors, previous cytogenetic and molecular cytogenetic data indicated that changes in chromosome 17 were associated with prostate carcinogenesis. Oshimura and Sandberg (1975) noted isochromosomes for 17q in a metastatic prostate tumor as early as 1975. In our laboratory, FISH analysis has demonstrated whole chromosome 17 loss in both cultured and archival prostate cancer specimens (Brothman et al., 1992, 1994; Jones et al., 1994).

There have been observations of an increased incidence of prostate cancer in families that show linkage for breast cancer. Arason and colleagues (1993) proposed that breast cancer genes may pre-

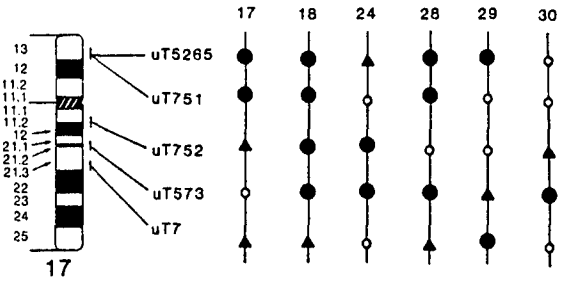


Figure 2. Schematic showing the location of the polymorphic markers (to the right of the ideogram) and regions lost in the six prostate tumors (UCAP 17, 18, 24, 28, 29, and 30). Solid circles represent allelic loss, open circles represent allelic retention, and triangles indicate that a marker was uninformative (homozygous at the locus studied).

dispose individuals to prostate cancer, based on a study of seven Icelandic families and association of 17q12-q23-linked markers. Furthermore, carriers of mutations in a gene predisposing to breast and ovarian cancer (*BRCA1*) have recently been shown to have an increased risk of colon and prostate cancer (Ford et al., 1994). Markers that we chose for this initial study included those that mapped near the *BRCA1* locus, because we sought to determine whether this locus was deleted in prostate cancers. In light of the recent isolation of the *BRCA1* gene (Miki et al., 1994), we have now determined that the most frequently lost marker, UT573, is within 1 Mb of *BRCA1*. Clearly, additional markers, including any that span *BRCA1*, must be evaluated.

Our findings suggest that a locus near *BRCA1* may be involved in early prostate cancer progression. Particularly interesting are specimens UCAP 24 and UCAP 30, which show interstitial deletions on the long arm of chromosome 17 (Fig. 2) in a region in close proximity to *BRCA1* (Miki et al., 1994; Williams et al., submitted). UCAP 28 also shows specific loss on 17q at UT573, but we have not yet determined whether this is interstitial, because the distal long-arm marker (UT7) was uninformative for this specimen. This specimen also showed loss of both short-arm markers. Likewise, both UCAP 17 and UCAP 29 showed specific loss of chromosome 17 short-arm material. Therefore, we suggest that other regions on chromosome 17, including portions of the short arm, may also be involved in prostate cancer. Previously, mutations in the *TP53* gene on 17p have been observed in advanced-stage prostate tumors (Macoska et al., 1992; Latil et al., 1994). Whether there are abnormalities in *TP53* or there is involvement of a different locus on 17p in our specimens remains to be seen.

One specimen in the current study, UCAP 18, appears to have lost an entire chromosome 17, because all informative markers on both sides of the centromere show deletions. Loss of one entire chromosome in this sample was also observed by FISH studies with pericentromeric probes (Jones et al., 1994) and a combination of pericentromeric and P1 probes (Williams et al., submitted). Whole-chromosome loss is an alternative mechanism that would reveal the presence of a tumor suppressor locus. UCAP 28 and UCAP 29 both yielded unusual results, in that there appears to be interstitial retention of chromosome 17 material (see Fig. 2). These two cases suggest that a complex rearrangement including the region near 17q21 has taken place; alternatively, two or more deletional events may have occurred.

The use of the PhosphorImager, in addition to the evaluation of replicate specimens, has proved invaluable in overcoming the problem of contaminating normal or stromal cells in our LOH study of prostate cancer. Microdissection of tumor tissue is a feasible alternative to multiple replicate analyses; yet, nontumor cells may persist in considerable amounts and cannot be eradicated completely. Bookstein and coworkers (1993) have very recently taken a similar approach to the analysis of prostate tumor specimens with PhosphorImager-analyzed data of markers on chromosome 8 but have examined a single measurement of each tumor/normal DNA pair for each specimen (Cher et al., 1994; MacGrogan et al., 1994). An advantage of limiting analyses to a single pair of DNAs is that many more markers can be readily evaluated, but this may be at the expense of sensitivity. The use of replicate samples increases the statistical power for evaluation and helps eliminate artifacts. The frequency of loss we observed at *LPL* (39% of informative cases) is comparable to that detected by MacGrogan et al. (1994), where 43% of informative cases showed LOH when a combination of three *LPL* polymorphic sites was used . It should also be noted that, although a limited number of markers has been evaluated at other chromosomal sites, we have not yet detected a high frequency of LOH on chromosomes other than 8 or 17 in prostate tumors using these methods.

Our method appears highly reliable and sensitive, having the ability to detect LOH in tumor cells with a relatively high background of normal cells. The mixing experiment (Table 1), in which normal DNA from one patient was added to a hybrid cell line containing one copy of chromosome 17 with an allele identical to that seen in the normal control, presents information about both the sensitivity and the limitations of the multiple-replicate PCR approach. Our data suggest that allelic

losses are detectable well above the traditionally accepted 50% level of contaminating normal cells (Table 1). Our analysis shows that this system has high sensitivity for LOH detection with only a moderate probability of missing LOH where it actually may be present. Thus, our detection rate of 27% of primary tumors with loss of chromosome 17 sequences is conservative.

Indeed, our parallel studies, in which we used FISH with selected bacteriophage P1 probes flanking the regions on 17q that showed loss, have confirmed that true loss exists in each specimen in which we detected LOH by PCR (Williams et al., submitted). Hence, by using two independent methods, we have indicated that regional loss on 17q occurs in prostate tumors, supporting the view that a tumor suppressor gene is present here.

The initial cellular processes that affect progression of prostate cancer are proving to be complex and may include multiple genetic events. Markers at additional chromosomal sites must be examined, so that the relationship between various genetic events in the progression of prostate tumorigenesis can be determined. The relatively small set of tumors examined in this study showed that approximately one-fourth of early prostate tumors have loss of at least some portion of chromosome 17, a frequency that compares favorably to observed losses on 10q and 16q (Carter et al., 1990). Our findings suggest that chromosome 17 is likely to play a significant role in this disease.

ACKNOWLEDGMENTS

The authors thank Drs. Rosie Plaetke and Ray White for their helpful discussions during the course of these studies. We also appreciate the efforts of individuals within the Utah Genome Center Marker Development and Mapping Group. This study was supported by NIH grants R01-CA46269, M01-0RR00064, and NCI CCSG 5P30 CA42014-08.

REFERENCES

Albertsen H, Plaetke R, Ballard L, Fujimoto E, Connolly J, Lawrence E, Rodriguez P, Robertson M, Bradley P, Miliner B, Fuhrman D, Marks A, Sargent R, Cartwright P, Matsunami N, White R (1994) Genetic mapping of the BRCA1 region on chromosome 17q21. Am J Hum Genet 54:516–525.

Arason A, Barkardottir RB, Egilsson V (1993) Linkage analysis of chromosome 17q markers and breast-ovarian cancer in Icelandic families, and possible relationship to prostatic cancer. Am J Hum Genet 52:711–717.

Bell GI, Karam JH, Rutter WJ (1981) Polymorphic DNA region adjacent to the 5′ end of the human insulin gene. Proc Natl Acad Sci USA 78:5759–5763.

Bergerheim USR, Kunimi K, Collins VP, Ekman P (1991) Deletion mapping of chromosomes 8, 10, and 16 in human prostatic carcinoma. Genes Chromosom Cancer 3:215–220.

Bookstein R, MacGrogan D, Milsenbeck SG, Sharkey F, Allred DC (1993) p53 is mutated in a subset of advanced-stage prostate cancers. Cancer Res 53:3369–3373.

Boring CC, Squires TS, Tong T, Montgomery S (1994) Cancer statistics. CA J Clin 44:7–26.

Bova GS, Carter BS, Bussemakers MJG, Emi M, Fujiwara Y, Natasha K, Jacobs SC, Robinson JC, Epstein JI, Walsh PC, Isaacs WB (1993) Homozygous deletion and frequent allelic loss of chromosome 8p22 loci in human prostate cancer. Cancer Res 53: 3869–3873.

Brewster SF, Browne S, Brown KW (1994) Somatic allelic loss at the DCC, APC, nm23-HI and p53 tumor suppressor gene loci in human prostatic carcinoma. J Urol 151:1073–1077.

Brothman AR, Patel AM, Peehl DM, Schellhammer PF (1992) Analysis of prostatic tumor cultures using fluorescence in situ hybridization (FISH). Cancer Genet Cytogenet 62:180–185.

Brothman AR, Watson MJ, Zhu XL, Williams BJ, Rohr LR (1994) Evaluation of 20 archival prostate tumor specimens by fluorescence in situ hybridization (FISH). Cancer Genet Cytogenet 75: 40–44.

Carter BS, Weing CM, Ward WS, Treiger BF, Aalders TW, Schalken JA, Epstein JI, Isaacs WB (1990) Allelic loss of chromosomes 16q and 10q in human prostate cancer. Proc Natl Acad Sci USA 87:8751–8755.

Cher ML, MacGrogan D, Bookstein R, Brown JA, Jenkins RB, Jensen RH (1994) Comparative genomic hybridization, allelic imbalance, and fluorescence in situ hybridization on chromosome 8 in prostate cancer. Genes Chromosom Cancer 11:153–162.

Effert PJ, Neubauer A, Walther PJ, Liu ET (1992) Alterations of the p53 gene are associated with the progression of human prostate carcinoma. J Urol 147:789–793.

Emi M, Yoshiyuki F, Nakajima T, Tsuchiya E, Tsuda H, Hirohashi S, Maeda Y, Tsuruta K, Miyaki M, Nakamura Y (1992) Frequent loss of heterozygosity for loci on chromosome 8p in hepatocellular carcinoma, colorectal cancer and lung cancer. Cancer Res 52: 5368–5372.

Emi M, Yoshiyuki F, Ohata H, Tsuda H, Hirohashi S, Koike M, Miyaki M, Mondne M, Nakamura Y (1993) Allelic loss at chromosome band 8p21.3-p22 is associated with progression of hepatocellular carcinoma. Genes Chromosom Cancer 7:152–157.

Ford D, Easton DF, Bishop DT, Narod SA, Goldgar DE (1994) Risks of cancer in *BRCA1* mutation carriers. Lancet 343:692–695.

Jones E, Zhu XL, Rohr LR, Stephenson RA, Brothman AR (1994) Aneusomy of chromosomes 7 and 17 detected by FISH in prostate cancer and the effects of in vitro selection. Genes Chromosom Cancer 11:163–170.

Kunimi K, Bergerheim USR, Larsson IL, Ekman P, Collins VP (1991) Allelotyping of human prostatic adenocarcinoma. Genomics 11:530–535.

Latil A, Baron JC, Cussenot O, Fournier G, Soussi T, Bacron-Gibod L, LeDuc A, Rovesse J, Lidereau R (1994) Genetic alterations in localized prostatic cancer: Identification of a common region of deletion on chromosome arm 18q. Genes Chromosom Cancer 11:119–125.

MacGrogan D, Levy A, Bostwick D, Wagner M, Wells D, Bookstein R (1994) Loss of chromosome 8p loci in prostate cancer: Mapping by quantitative allelic imbalance. Genes Chromosom Cancer 10:151–159.

Macoska JA, Powell IJ, Sakr W, Lane MA (1992) Loss of the 17p chromosomal region in a metastatic carcinoma of the prostate. J Urol 147:1142–1146.

Macoska JA, Micale MA, Sakr WA, Benson PD, Wolman SR (1993) Extensive genetic alterations in prostate cancer revealed by dual PCR and FISH analysis. Genes Chromosom Cancer 8:88–97.

Miki Y, Swensen J, Shattuck-Eldens D, Futreal PA, Harshman K, Tavtigian S, Qingyun L, Cochran C, Bennett LM, Ding W, Bell R, Rosenthal J, Hussey C, Tran T, McClure M, Frye C, Hattier T, Phelps R, Haugen-Strano A, Katcher H, Yakumo K, Gholami Z, Shaffer D, Stone S, Bayer S, Wray C, Bogden R, Dayananth P, Ward J, Tonin P, Narod S, Bristow PK, Norris FH, Helvering H, Morrison P, Rosteck P, Lai M, Barrett JC, Lewis C, Neuhausen S, Cannon-Albright L, Goldgar D, Wiseman R, Kamb A, Skolnick MH (1994) A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science 266:66–71.

Oshimura M, Sandberg AA (1975) Isochromosome 17 in prostate cancer. J Urol 114:249–250.

Utah Marker Development Group (in press) A collection of ordered tetra-nucleotide repeat markers from the human genome. Am J Hum Genet.

Zuliani G, Hobbs HH (1990) Tetranucleotide repeat polymorphism in the LPL gene. Nucleic Acids Res 18:4958.

GENOMICS 28, 25–31 (1995)

A Gene (*DLG2*) Located at 17q12–q21 Encodes a New Homologue of the Drosophila Tumor Suppressor dlg-A

SYLVIE MAZOYER,[1] SIMON A. GAYTHER, MARIA A. NAGAI,[2] SIMON A. SMITH,* ALISON DUNNING, ELIZABETH J. VAN RENSBURG,[3] HANS ALBERTSEN,* RAY WHITE,* AND BRUCE A. J. PONDER

*CRC Human Cancer Genetics Research Group, Laboratories Block, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2QP, United Kingdom; and *Eccles Institute of Human Genetics and Howard Hughes Medical Institute, University of Utah, Salt Lake City, Utah 84112*

Received December 13, 1994; accepted April 17, 1995

We have isolated a novel cDNA that maps distal to BRCA1 at 17q12–q21. The total sequence predicts a protein of 576 amino acids with three conserved regions: a 90-amino-acid repeat domain, a SH3 (*src* homology region 3) motif, and a guanylate kinase domain. These conserved regions are shared among members of the discs-large family of proteins that include human p55, a membrane protein expressed in erythrocytes, rat PSD-95/SAP90, a synapse protein expressed in brain, Drosophila dlg-A, a septate junction protein expressed in various epithelia, and human and mouse ZO-1 and canine ZO-2, two tight junction proteins. dlg-A has been shown to act as a tumor suppressor, and the other members may all be involved in signal transduction through specialized membrane domains with highly organized cytoskeletons and thus are potential tumor suppressors. Since allelic loss has been reported in the 17q12–q21 region in breast and ovarian cancer and it appears that BRCA1 is not the target of the losses, we looked for somatic alterations in *DLG2* in sporadic breast tumors. No evidence for mutation was found, making it unlikely that *DLG2* is involved in sporadic breast cancer.

INTRODUCTION

While screening human cDNA libraries with reagents used to generate a 4-cM physical map in 17q12–q21 (Albertsen *et al.,* 1994), we have identified a new gene encoding a protein that contains a 90-amino-acid repeat domain of unknown function, a SH3 (*src* homology region 3) motif, and a guanylate kinase domain and thus belongs to the discs-large family of proteins. This family comprises the Drosophila *discs-large* tumor suppressor product, dlg-A, localized to the septate junctions in various epithelia (Woods and Bryant, 1991), the human p55 membrane protein expressed in erythrocytes (Ruff *et al.,* 1991), PSD-95/SAP90, a presynaptic junction protein from rat brain (Kistner *et al.,* 1993), and human and mouse ZO-1 and dog ZO-2, two tight junction proteins (Willot *et al.,* 1993; Itoh *et al.,* 1993; Jesaitis and Goodenough, 1994). As a result of this homology, we named this new gene *DLG2*.

Allelic loss on chromosome arm 17q has been reported in sporadic breast and ovarian tumors (Jacobs *et al.,* 1993; Saito *et al.,* 1993; Nagai *et al.,* 1994), and detailed maps delimiting the area of losses have been constructed to investigate whether the breast and ovarian cancer susceptibility gene BRCA1 located in 17q12–q21 is involved in the genesis of sporadic as well as inherited tumors, as is usually the case with genes predisposing to cancer. BRCA1 has been recently cloned (Miki *et al.,* 1994). Contrary to expectations, no somatic mutations were identified in 44 sporadic breast and ovarian tumors studied, suggesting that BRCA1 is not involved in noninherited forms of the disease (Futreal *et al.,* 1994). However, because LOH are frequent in or adjacent to the BRCA1 region (Nagai *et al.,* 1994; Cropp *et al.,* 1994; Futreal *et al.,* 1994), one or even more tumor suppressor genes may be located in the same region (Ponder, 1994; Vogelstein and Kinzler, 1994). As the *DLG2* gene was a good candidate, we looked for rearrangements in *DLG2* in DNA from sporadic breast and ovarian cancers and for mutations in a set of sporadic breast tumors for which genomic DNA was also available and that showed LOH in 17q.

MATERIALS AND METHODS

DLG2 cDNA isolation and sequencing. A cDNA clone, 38B1/1, containing a 3.1-kb insert was isolated using a combination of 3 P1

Sequence data from this article have been deposited with the EMBL Data Library under Accession No. X82895.

[1] To whom correspondence should be addressed. Telephone: 44 1223 336923. Fax: 44 1223 336902. E-mail: smazoyer@hgmp.mrc.ac.uk.

[2] Present address: Disciplina de Oncologia, Departamento de Radiologia da Faculdade de Medicina da Universidade de São Paulo, São Paulo, Brazil.

[3] Present address: Department of Human Genetics and Developmental Biology, Medical Faculty of the University of Pretoria, Pretoria, South Africa.

0888-7543/95 $12.00

phages, 146B3, 750G12, and 92E12 when screening a fetal brain cDNA library (Stratagene, Cat. No. 936206). Prior to further analysis, this cDNA was shown to map back to 17q12–q21 when hybridized to a filter containing P1 phages and YACs DNA covering the BRCA1 region (see Fig. 2, Albertsen *et al.*, 1994). cDNA sequence was then determined using a *Taq* DyeDeoxy Terminator Cycle sequencing kit on a automated sequencer ABI 373A (Applied Biosystems). The BLAST algorithm was used to look for homologies or identities between the sequence identified and other sequences from databases (GenBank/EMBL at the nucleotide level, Swissprot/NBRF at the amino acid level). While a polyadenylation signal and the beginning of a poly(A) tail were present in 38B1/1, the 5′ end of the cDNA appeared to be missing. We then used a 5′-RACE-Ready testis cDNA kit (Clontech) to isolate additional 5′ region. The PCR fragment generated using this kit was cloned in a TA cloning vector (InVitrogen) and sequenced as described. Exon–intron boundaries were localized as follows: primers distributed evenly along the cDNA sequence were used to amplify fragments from genomic DNA as well as from the 38B1/1 clone. Whenever a difference in size was observed between genomic DNA and cDNA, the genomic fragment was subcloned in a TA cloning vector and sequenced and intron–exon boundaries were determined by comparison with the cDNA sequence. Because we were unable to amplify from genomic DNA in the very 5′ end of the coding sequence, certainly due to the presence of a big intron, the structure of this part of the gene has not been completely resolved.

Northern and Southern blots. A multiple human tissue Northern and a zoo blot were purchased from Clontech and hybridized as recommended. Southern filters were made from 5 μg *Eco*RI- or *Rsa*I-digested DNA purified from paired 28 blood/sporadic breast tumors and 33 blood/sporadic ovarian tumors. Hybridization and washes were performed using standard protocols. Probes were labeled using the random-primed labeling protocol.

Sequence comparison analysis. Protein sequences were compared by the PILEUP and GAP algorithms using the UK Human Genome Mapping Project computing services.

Mutation screening. Paired primary breast tumors and normal tissue DNA samples were obtained from patients at the A. C. Camargo Hospital, São Paulo, Brazil, and were fully described elsewhere (Nagai *et al.*, 1994).

One hundred nanograms of DNA was amplified in a 50-μl reaction using the following primers (see Fig. 4): 9F, GAAGAGACCCGGGACACTG; 9R, ATTCTCAATCCCCCCACC; 10F, ATTCTGACTTGGAGCAATTGG; 10R, GACAGCCAGGAGGCTCAG; 1F, AGCCTTCTGCTGACCCTTC; 1R, CCCTCCCATGCACCATAC; 2F, GTATGGTGCATGGGAGGG; 11R, AAGGCATGCCATGTTAGAGG; 11F, AGCCTCCCTGAGAGCTCC; 3R, GAGCCCTGCTCCTTGTCC; 4F, GAGGCACTGTAACCTGCCC; 4R, AGGCAGCAGAGAGGACATTG; 5F, TCCTAGGACAGACATGGGGA; 5R, CCACCCTAGGCAGCTATCAG; 6F, CTTCTGACAGTGGGGGTGAG; 6R, TCAGGTTCTGGGTTTCAACA. Exons 7 and 8 were amplified together because of their small size and because of the intron between them, whereas the other exons were amplified separately.

PCR products were diluted with an equal volume of 95% formamide and denatured at 100°C for 10 min before quenching on ice. Prior to loading and polyacrylamide gel electrophoresis, samples were retained on ice for 10 min to allow a proportion of the single-strand DNA to reanneal, thereby encouraging the formation of any potential double-stranded heteroduplexes. Electrophoresis was performed in nondenaturing 0.5× MDE polyacrylamide gels (J. T. Baker) using a Protean II vertical gel apparatus (Bio-Rad). Twenty-centimeter gels were run at 200 V for 8–12 h and the SSCP conditions optimized for temperatures between 4 and 15°C. Polyacrylamide gels were stained with 0.1% silver nitrate solution for instantaneous visualization of results.

RESULTS

Isolation of DLG2

We isolated a 3.1-kb cDNA from a human fetal brain cDNA library using a mixture of three P1 phages localized between D17S78 and 17HSD on chromosome band 17q12–q21 (Albertsen *et al.*, 1994). Comparison of the sequence of this cDNA clone, called 38B1/1, to those in the GenBank/EMBL database failed to reveal any convincing homology to known genes. However, the amino acid sequence deduced from the longest open reading frame showed a strong homology with p55 (discussed below), which is a major palmitoylated membrane protein of human erythrocytes (Ruff *et al.*, 1991), and to a lesser extent with a Drosophila tumor suppressor protein, dlg-A (Woods and Bryant, 1991) and with a wide range of guanylate kinases (GK), the best homology being with pig GK. This gene was called *DLG2*. Since the initiation codon was not present in 38B1/1, the 5′ end of *DLG2* was then pursued using the 5′ RACE technique, which allowed us to identify an additional 357 bp. The nucleotide sequence and predicted amino acid sequence of the full-length cDNA are shown in Fig. 1. The first methionine codon is at nucleotide 88 and reveals an open reading frame of 1731 bp. If this ATG represents the initiation codon as suggested by the presence of a purine 3 bp downstream, the open reading frame codes for a 576-amino-acid protein of a predicted molecular weight of 64.5 kDa. The presence of an in-frame stop codon upstream from the initiation codon further supports that the sequence coding for the N-terminal end of the protein has been isolated. The 3′ untranslated region is 1651 bp long and contains a consensus polyadenylation signal, AATAAA, followed by the beginning of a poly(A) tail 15 bp downstream. Two RNA species of approximately 3.8 and 4.8 kb were revealed on a Northern blot (Fig. 2a); the total sequence isolated is consistent with the size of the smaller mRNA. Attempts to isolate further sequence in the 5′ end using the 5′ RACE technique was unsuccessful. Alternative splicing was observed in the 5′ end (data not shown) but the splices identified cannot account for the 4.8-kb mRNA.

Northern Blot and Zoo Blot Analysis

To assess the pattern of expression of dlg2 mRNA in normal tissue, the 38B1/1 insert was used as a probe in a Northern blot assay (Fig. 2a). Two mRNA of 4.8 and 3.8 kb were detected in prostate, testis, ovary, small intestine, and colon but not in spleen, thymus, and peripheral blood leukocytes. The level of expression of these mRNAs is higher in testis than in the other tissues, while hybridization with an actin probe indicated that each lane contained equal amounts of RNA (data not shown). The *DLG2* gene was also found to be expressed in breast epithelium and placenta when cDNA made from total RNA extracted from these two tissues was used as template in PCR experiments (data not shown).

Zoo blot analysis using the 38B1/1 insert as a probe demonstrated sequence conservation in monkey, rat, mouse, dog, and cow, but not rabbit, chicken, or yeast (Fig. 2b).

```
                    GGAGCGC  CCGGCTGCGC  TGGAGCCGCC  CGGAGCTAGG  GGCTCCCCGG  GGCGCAGGAG  AGACGTTTCA  GAGCCCTTGC  CTCCTTCACC 87

ATG CCG GTT GCC GCC ACC AAC TCT GAA ACT GCC ATG CAG CAA GTC CTG GAC AAC TTG GGA TCC CTC CCC AGT GCC ACG GGG GCT GCA GAG 177
Met pro val ala ala thr asn ser glu thr ala met gln gln val leu asp asn leu gly ser leu pro ser ala thr gly ala ala glu 30

                                                                                                    ▽
CTG GAC CTG ATC TTC CTT CGA GGC ATT ATG GAA AGT CCC ATA GTA AGA TCC CTG GCC AAG GTG ATA ATG GTA TTG TGG TTT ATG CAG CAG 267
leu asp leu ile phe leu arg gly ile met glu ser pro ile val arg ser leu ala lys val ile met val leu trp phe met gln gln 60

                                                    ▽
AAT GTC TTT GTT CCT ATG AAA TAC ATG CTG AAA TAC TTT GGG GCC CAT GAG AGG CTG GAG GAG ACG AAG CTG GAG GCC GTG AGA GAC AAC 357
asn val phe val pro met lys tyr met leu lys tyr phe gly ala his glu arg leu glu glu thr lys leu glu ala val arg asp asn 90

.
AAC CTG GAG CTG GTG CAG GAG ATC CTG CGG GAC CTG GCG CAG CTG GCT GAG CAG AGC AGC ACA GCC GCC GAG CTG GCC CAC ATC CTC CAG 447
asn leu glu leu val gln glu ile leu arg asp leu ala gln leu ala glu gln ser ser thr ala ala glu leu ala his ile leu gln 120

          ▽
GAG CCC CAC TTC CAG TCC CTC CTG GAG ACG CAC GAC TCT GTG GCC TCA AAG ACC TAT GAG ACA CCA CCC CCC AGC CCT GGC CTG GAC CCT 537
glu pro his phe gln ser leu leu glu thr his asp ser val ala ser lys thr tyr glu thr pro pro pro ser pro gly leu asp pro 150

                                                                                                    ▽
ACG TTC AGC AAC CAG CCT GTA CCT CCC GAT GCT GTG CGC ATG GTG GGC ATC CGC AAG ACA GCC GGA GAA CAT CTG GGT GTA ACG TTC CGC 627
thr phe ser asn gln pro val pro pro asp ala val arg met val gly ile arg lys thr ala gly glu his leu gly val thr phe arg 180

GTG GAG GGC GGC GAG CTG GTG ATC GCG CGC ATT CTG CAT GGG GGC ATG GTG GCT CAG CAA GGC CTG CTG CAT GTG GGT GAC ATC ATC AAG 717
val glu gly gly glu leu val ile ala arg ile leu his gly gly met val ala gln gln gly leu leu his val gly asp ile ile lys 210
                  ++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++

GAG GTG AAC GGG CAG CCA GTG GGC AGT GAC CCC CGC GCA CTG CAG GAG CTC CTG CGC AAT GCC AGT GGC AGT GTC ATC CTC AAG ATC CTG 807
glu val asn gly gln pro val gly ser asp pro arg ala leu gln glu leu leu arg asn ala ser gly ser val ile leu lys ile leu 240
++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++++

                                ▽
CCC AAC TAC CAG GAG CCC CAT CTG CCC CGC CAG GTA TTT GTG AAA TGT CAC TTT GAC TAT GAC CCG GCC CGA GAC AGC CTC ATC CCC TGC 897
pro asn tyr gln glu pro his leu pro arg gln val phe val lys cys his phe asp tyr asp pro ala arg asp ser leu ile pro cys 270
                                ==============================================================================

                                                                                                    ▽
AAG GAA GCA GGC CTG CGC TTC AAC GCC GGG GAC TTG CTC CAG ATC GTA AAC CAG GAT GAT GCC AAC TGG TGG CAG GCA TGC CAT GTC GAA 987
lys glu ala gly leu arg phe asn ala gly asp leu leu gln ile val asn gln asp asp ala asn trp trp gln ala cys his val glu 300
================================================================================================================================

GGG GGC AGT GCT GGG CTC ATT CCC AGC CAG CTG CTG GAG GAG AAG CGG AAA GCA TTT GTC AAG AGG GAC CTG GAG CTG ACA CCA AAC TCA 1077
gly gly ser ala gly leu ile pro ser gln leu leu glu glu lys arg lys ala phe val lys arg asp leu glu leu thr pro asn ser 330
==========================================================

▽                                                                                                   ▽
GGG ACC CTA TGC GGC AGC CTT TCA GGA AAG AAA AAG AAG CGA ATG ATG TAT TTG ACC ACC AAG AAT GCA GAG TTT GAC CGT CAT GAG CTG 1167
gly thr leu cys gly ser leu ser gly lys lys lys lys arg met met tyr leu thr thr lys asn ala glu phe asp arg his glu leu 360

CTC ATT TAT GAG GAG GTG GCC CGC ATG CCC CCG TTC CGC CGG AAA ACC CTG GTA CTG ATT GGG GCT CAG GGC GTG GGA CGG CGC AGC CTG 1257
leu ile tyr glu glu val ala arg met pro pro phe arg arg lys thr leu val leu ile gly ala gln gly val gly arg arg ser leu 390

                                                    ▽
AAG AAC AAG CTC ATC ATG TGG GAT CCA GAT CGC TAT GGC ACC ACG GTG CCC TAC ACC TCC CGG CGG CCG AAA GAC TCA GAG CGG GAA GGT 1347
lys asn lys leu ile met trp asp pro asp arg tyr gly thr thr val pro tyr thr ser arg arg pro lys asp ser glu arg glu gly 420

CAG GGT TAC AGC TTT GTG TCC CGT GGG GAG ATG GAG GCT GAC GTC CGT GCT GGG CGC TAC CTG GAG CAT GGC GAA TAC GAG GGC AAC CTG 1437
gln gly tyr ser phe val ser arg gly glu met glu ala asp val arg ala gly arg tyr leu glu his gly glu tyr glu gly asn leu 450

                                                                                                    ▽
TAT GGC ACA CGT ATT GAC TCC ATC CGG GGC GTG GTC GCT GCT GGG AAG GTG TGC GTG CTG GAT GTC AAC CCC CAG GCG GTG AAG GTG CTA 1527
tyr gly thr arg ile asp ser ile arg gly val val ala ala gly lys val cys val leu asp val asn pro gln ala val lys val leu 480

CGA ACG GCC GAG TTT GTC CCT TAC GTG GTG TTC ATC GAG GCC CCA GAC TTC GAG ACC CTG CGG GCC ATG AAC AGG GCT GCG CTG GAG AGT 1617
arg thr ala glu phe val pro tyr val val phe ile glu ala pro asp phe glu thr leu arg ala met asn arg ala ala leu glu ser 510

                                ▽
GGA ATA TCC ACC AAG CAG CTC ACG GAG GCG GAC CTG AGA CGG ACA GTG GAG GAG AGC AGC CGC ATC CAG CGG GGC TAC GGG CAC TAC TTT 1707
gly ile ser thr lys gln leu thr glu ala asp leu arg arg thr val glu glu ser ser arg ile gln arg gly tyr gly his tyr phe 540

GAC CTC TGC CTG GTC AAT AGC AAC CTG GAG AGG ACC TTC CGC GAG CTC CAG ACA GCC ATG GAG AAG CTA CGG ACA GAG CCC CAG TGG GTG 570
asp leu cys leu val asn ser asn leu glu arg thr phe arg glu leu gln thr ala met glu lys leu arg thr glu pro gln trp val 1797

CCT GTC AGC TGG GTG TAC TGA                                                                                                1818
pro val ser trp val tyr OPA                                                                                                576

 GCCTGTTCAC  CTGGTCCTTG  GCTCACTCTG  TGTTGAAACC  CAGAACCTGA  ATCCATCCCC  CTCCTGACCT  GTGACCCCCT  GCCACAATCC  TTAGCCCCCA 1918
 TATCTGGCTG  TCCTTGGGTA  ACAGCTCCCA  GCAGGCCCTA  AGTCTGGCTT  CAGCACAGAG  GCGTGCACTG  CCAGGGAGGT  GGGCATTCAT  GGGGTACCTT 2018
 GTGCCCAGGT  GCTGCCCACT  CCTGATGCCC  ATTGGTCACC  AGATATCTCT  GAGGGCCAAG  CTATGCCCAG  GAATGTGTCA  GAGTCACCTC  CATAATGGTC 2118
 ACAGAGAAGA  AGTGAAAAGC  TGCTTTGGGA  CCACATGGTC  AGTAGGCACA  CTGCCCCCTG  CCACCCCTCC  CCAGTCACCA  GTTCTCCTCT  GGACTGGCCA 2218
 CACCCACCCC  ATTCCTGGAC  TCCTCCCACC  TCTCACCTCT  GTGTCGGAGG  AACAGGCCTT  GGGCTGTTTC  CGTGTGACCA  GGGGAATGTG  TGGCCCGCTG 2318
 GCAGCCAGGC  AGGCCCGGGT  GGTGGTGCCA  GCCTGGTGCC  ATCTTGAAGG  CTGGAGGAGT  CAGAGTGAGA  GCCAGTGGCC  ACAGCTGCAG  AGCACTGCAG 2418
 CTCCCAGCTC  CTTTGGAAAG  GGACAGGGTC  GCAGGGCAGA  TGCTGCTCGG  TCCTTCCCTC  ATCCACAGCT  TCTCACTGCC  GAAGTTTCTC  CAGATTTCTC 2518
 CAATGTGTCC  TGACAGGTCA  GCCCTGCTCC  CCACAGGGCC  AGGCTGGCAG  GGGCCAGTGG  GTTCAGCCCA  GGTAGGGGCA  GGATGGAGGG  CTGAGCCCTG 2618
 TGACAACCTG  CTGTTACCAA  CTGAAGAGCC  CCAAGCTCTC  CATGGCCCAC  AGCAGGCACA  GGTCTGAGCT  CTATGTCCTT  GACCTTGGTC  CATTTGGTTT 2718
 TCTGTCTAGC  CAGGTCCAGG  TAGCCCACTT  GCATCAGGGC  TGCTGGGTTG  GAGGGGCTAA  GGAGGAGTGC  AGAGGGGACC  TTGGGAGCCT  GGGCTTGAAG 2818
 GACAGTTGCC  CTCCAGGAGG  TTCCTCACAC  ACAACTCCAG  AGGCGCCATT  TACACTGTAG  TCTGTACAAC  CTGTGGTTCC  ACGTGCATGT  TCGGCACCTG 2918
 TCTGTGCCTC  TGGCACCAGG  TTGTGTGTGT  GTGCGTGTGC  ACGTGCGTGT  GTGTGTGTGT  GTGTCAGGTT  TAGTTTGGGG  AGGAACCAAA  GGGTTTTGTT 3018
 TTGGAGGTCA  CTCTTTGGGG  CCCCTTTCTG  GGGGTTCCCC  ATCAGCCCTC  ATTTCTTATA  ATACCCTGAT  CCCAGACTCC  AAAGCCCTGG  TCCTTTCCTG 3118
 ATGTCTCCTC  CCTTGTCTTA  TTGTCCCCCT  ACCCTAAATG  CCCCCCTGCC  ATAACTTGGG  GAGGGCAGTT  TTGTAAAATA  GGAGACTCCC  TTTAAGAAAG 3218
 AATGCTGTCC  TAGATGTACT  TGGGCATCTC  ATCCTTCATT  ATTCTCTGCA  TTCCTTCCGG  GGGGAGCCTG  TCCTCAGAGG  GGACAACCTG  TGACACCCTG 3318
 AGTCCAAACC  CTTGTGCCTC  CCAGTTCTTC  CAAGTGTCTA  ACTAGTCTTC  GCTGCAGCGT  CAGCCAAAGC  TGGCCCCTGA  ACCACTGTGT  GCCCATTTCC 3418
 TAGGGAAGGG  GAAGGAGAAT  AAACAGAATA  TTTATTACAA  AAA                                                                    3461
                         ***  ***
```

FIG. 1. Nucleotide sequence of *DLG2*. The nucleotide sequence of *DLG2* is presented along with the predicted amino acid sequence. The positions of 11 introns are indicated with arrowheads above the sequence, the DHR domain is underlined with crosses, the SH3 domain is double-underlined, and the guanylate domain is single-underlined. Poly(A) signal is underlined with asterisks. The first nucleotide of clone 38B1/1 is indicated with a dot.

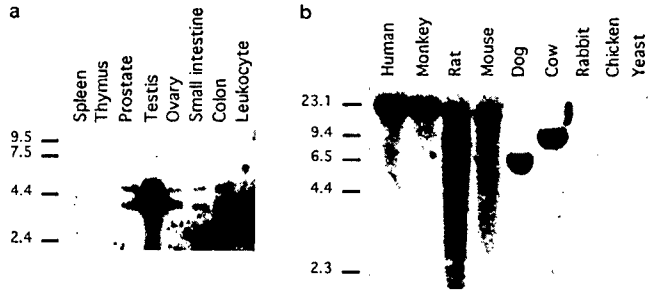


FIG. 2. (**a**) Northern analysis of *DLG2*. A human multiple-tissue Northern blot was probed with the 38B1/1 insert. Two mRNA species of 4.8 and 3.8 kb are detected in prostate, testis, ovary, small intestine, and colon but not in spleen, thymus, and peripheral blood leukocytes. Size standards are in kilobases. (**b**) Conservation analysis of *DLG2*. An *Eco*RI zoo blot was probed with the 38B1/1 insert. Fragments are detected in monkey, rat, mouse, dog, and cow, but not in rabbit, chicken, or yeast. Size standards are in kilobases.

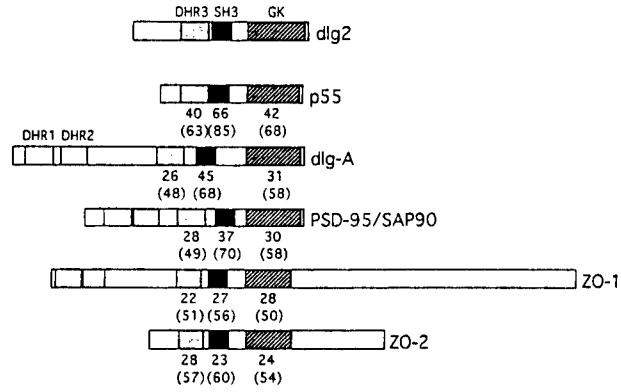


FIG. 3. Schematic comparison of structural domains of dlg2 with those of p55, dlg-A, PSD-95/SAP90, ZO1, and ZO2 (DHR, 90-aa internal repeat; SH3, *src* homology 3; GK, guanylate kinase domain). Percentage identity and percentage similarity (in parentheses) for each domain were calculated using the GAP algorithm.

dlg2 Belongs to the Discs-Large Family of Proteins

The predicted amino acid sequence of dlg2 is homologous to the human erythrocyte membrane protein p55 (Ruff *et al.,* 1991) (42% identity, 64% similarity), the rat brain synapses protein PSD-95/SAP90 (Kistner *et al.,* 1993) (27% identity, 53% similarity), the Drosophila dlg-A tumor suppressor (Woods and Bryant, 1991) (26% identity, 48% similarity), the dog tight-junction protein ZO-2 (Jesaitis and Goodenough, 1994) (22% identity, 50% similarity) and the human tight-junction ZO-1 (Willott *et al.,* 1993) (20% identity, 48% similarity) (Table 1). The sequence homologies of these family members are clustered in three domains (Fig. 3): a 90-amino-acid internal repeat region called DHR (discs-large homology region) (Bryant *et al.,* 1993), a SH3 domain (*src* homology 3) which was first identified in the noncatalytic region of the src family of protein tyrosine kinases (Musacchio *et al.,* 1992), and now found in multiple signaling and cytoskeletal proteins (Mayer and Baltimore, 1993), and a guanylate kinase domain.

While ZO-1, PSD-95/SAP90, and dlg-A have three DHR domains, DHR1, DHR2, and DHR3, p55, and dlg2 contain only one repeat (Fig. 3). Although this repeat shows greater homology to the third repeat of dlg-A, this homology is not very high (26% identity, 48% similarity for dlg2, 28% identity, 48% similarity for p55), while it shows 40% identity (63% similarity) between dlg2 and p55. The N-terminal amino acid sequence of ZO-2 has not yet been deduced, and although only one repeat, DHR3, has been found so far, two more domains may be identified, as suggested by the high degree of homology between ZO-1 and ZO-2 (Jesaitis and Goodenough, 1994). The SH3 domain of dlg2 is more homologous to the SH3 domain of p55 than to that of dlg-A, PSD-95/SAP90, ZO-1, and ZO-2 (Fig. 3). Also, the dlg2 guanylate kinase domain shows closer homology to that of p55 than that of the other members of the discs-large family. Furthermore, while three of eight amino acids that are thought to bind the ATP phosphate donor (Stehle and Schulz, 1990) are deleted in the guanylate kinase domain of dlg-A, PSD-95/SAP90, ZO-1, and ZO-2 (Koonin *et al.,* 1992; Willott *et al.,* 1993), both dlg2 and p55 retain these residues. Nevertheless, this motif (GxxGxGKS/T) does show substitutions in highly conserved residues in both proteins. These might affect enzyme activity (GxxGxGRR for dlg-1; GxxGxRS for p55); however, the Lys to Arg substitution has been observed in several putative ATPases (Gorbalenya and Koonin, 1990).

TABLE 1

Discs-Large Family Member Comparisons

p55	PSD-95/ SAP90	dlg-A	ZO-1	ZO-2	
42 (64)	27 (53)	26 (48)	20 (48)	22 (50)	dlg2
	27 (55)	28 (52)	24 (47)	20 (46)	p55
		58 (76)	31 (55)	17 (41)	PSD-95/ SAP90
			26 (47)	20 (42)	dlg-A
				52 (69)	ZO-1

Note. Percentage identity and percentage similarity (in brackets) were calculated for each pair throughout the protein using the GAP algorithm. The highest scores are shown in bold type.

Mutation Analysis of DLG2

To determine whether *DLG2* is the target for the allele losses frequently reported in the BRCA1 region in sporadic breast and ovarian cancer (Jacobs *et al.,* 1993; Saito *et al.,* 1993; Nagai *et al.,* 1994; Cropp *et al.,* 1994), we first hybridized the major part of dlg2 cDNA with a set of Southern filters containing DNA purified from blood/tumor pairs from 28 sporadic breast and 33 sporadic ovarian tumors. No rearrangements were found. We then performed a more sensitive mutation analysis by screening genomic and sporadic breast tumor DNA from 19 patients using single-strand conformation polymorphism (SSCP) and heteroduplex analyses. This set of samples was chosen on the basis of displaying losses on 17q, as has been described else-

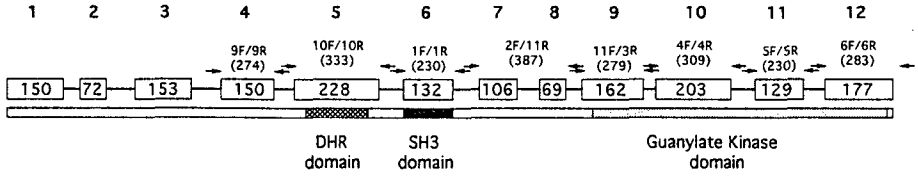


FIG. 4. Schematic structure of the *DLG2* gene. Intron–exon structure is shown, as well as primers used for the mutation analyses. The size of each exon is given in the box, while the sizes of the PCR fragments are given in parentheses; the sizes of the introns are not to scale. The conserved domains are shown below.

where (Nagai *et al.*, 1994), for D17S855, which is located in an intron of the BRCA1 gene (Futreal *et al.*, 1994) and is close to *DLG2*. Exons covering 78% of the coding sequence including the DHR, SH3, and guanylate kinase domains were amplified using primers located in adjacent introns (Fig. 4). Single-strand variants were observed in 10 tumors, some of which displayed abnormal fragments in more than one exon, but because these fragments were present in germ-line DNA as well, they were assumed to represent polymorphisms. A total of five different variants were found in four different exons (representative SSCP and heteroduplex analyses are shown in Fig. 5). Eight individuals were heterozygous in exon 4, three in exon 11, and one in exon 12. In exon 5, two different types of variants were found, three individuals being heterozygous for the first variation and four for the second one. However, no aberrant bands were observed in the tumor samples only and not in blood. Both the Southern and the SSCP/heteroduplex analyses therefore suggest that *DLG2* is not mutated in sporadic breast cancers.

DISCUSSION

One of the clones identified from screening a human fetal brain library, which mapped to 17q12–q21, showed a strong homology at the amino acid level with an erythroid membrane protein, p55 (Ruff *et al.*, 1991), and to a lesser extent with a Drosophila septate junction protein, dlg-A (Woods and Bryant, 1991), a rat presynaptic protein, PSD-95/SAP90 (Kistner *et al.*, 1993), and two tight junction proteins, human (Willott *et al.*, 1993) and mouse ZO-1 (Itoh *et al.*, 1993) and canine ZO-2 (Jesaitis and Goodenough, 1994). This protein, which we named dlg2, contains 576 amino acids and displays all of the features shared by proteins that belong to the discs-large family. These include a 90-amino-acid motif of unknown function called DHR, an SH3 domain, and a guanylate kinase domain. The DHR domain, also present in nitric oxide synthetase (Cho *et al.*, 1992), a human cytosolic tyrosine phosphatase (Gu *et al.*, 1991), and the cytoplasmic domain of a human brain transmembrane protein (Duclos *et al.*, 1993) might be involved in directing membrane association. The SH3 domain has been shown to be involved in the targeting of proteins to specific subcellular locations (Bar-Sagi *et al.*, 1993). It is also possible that certain SH3 domains participate in the control of cytoskeletal organization. The yeast guanylate kinase (GK) catalyzes the conversion of GMP to GDP by transferring a phosphate group from ATP. Although the GK domains of dlg2 and p55 retain the conserved amino acids thought to interact with ATP and thus are more likely to exert an enzymatic activity than those of ZO-1, ZO-2, dlg-A, and PSD-95/SAP90, direct evidence for such an activity is needed. An alternative function suggested for these GK domains comes from the observation that the structure of yeast GK is similar to that of small GTP-binding proteins (Stehle and Schulz, 1990). Thus, members of the discs-large family of proteins could be involved in signaling pathways through interaction with G-protein-binding proteins and not by altering GDP/GTP ratios.

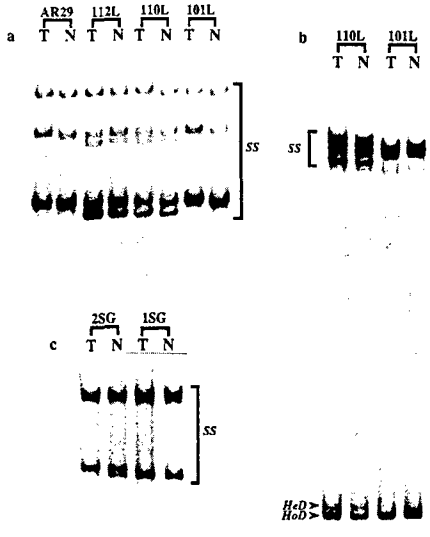


FIG. 5. Mutation analysis of *DLG2* in breast tumors. Representative pattern of single-strand (SS) and double-stranded hetero- (HeD) and homo- (HoD) duplexes for paired tumor (T) and normal (N) tissue of individuals for (**a**) exon 4—samples 112L and 110L show SS variants; (**b**) exon 5—sample 110L demonstrates variation in both SS and HeD patterns; and (**c**) exon 11—sample 2SG demonstrates SS variation.

The dlg-A, ZO-1, ZO-2, and PSD-95/SAP90 proteins are localized to regions of cell–cell contact, and p55 is copurified during the isolation of dematin, an actin-bundling protein of the erythrocyte membrane cytoskeleton (Ruff *et al.*, 1991); dlg2 might then be localized at the membrane of the cells in which it is expressed, possibly more specifically in specialized membrane domains with a highly organized cytoskeleton.

Thus far, only two tissues have been tested for the expression of both dlg2 and p55: while p55 alone is

expressed in red cells, both are expressed in placenta (Metzenberg and Gitschier, 1992). The coexpression of these two very closely related proteins implies a specific function for the first 81 amino acids of dlg2, the only part of this protein not homologous to p55. It should be noted that all discs-large protein encoding genes reported to date, except ZO-1 and ZO-2, give rise to at least two mRNAs both expressed in the same tissues, although at different levels (Woods and Bryant, 1991; Ruff *et al.,* 1991; Kistner *et al.,* 1993). Nothing is known about how these two mRNA species are generated. In the case of PSD-95/SAP90, the larger mRNA species disappears during development (Kistner *et al.,* 1993). Only one of the corresponding sequences is available for each of these four genes, and it would be interesting to know whether these isoforms show any homology to or affect the specificity of each of these proteins. It is also noteworthy that all of these genes are subject to alternative splicing.

The first member of this protein family to be cloned is also the one whose function is best elucidated. Recessive lethal mutations in the Drosophila *dlg* gene produce a neoplastic overgrowth phenotype in imaginal discs, due to an absence of proliferation control, apical–basal cell polarity, cell adhesion, and ability of cells to differentiate (Woods and Bryant, 1991). Mutation analysis indicates that both the GK domain function and the amino-terminal end of the gene product are important for its tumor suppressing function (Woods and Bryant, 1991). The GK domain of dlg-A, even if not functional as such, is potentially involved in guanine nucleotide metabolism or regulation, which suggests a relationship with signal transduction mechanisms known to control cell proliferation. Because of their homology to dlg-A, all of the members of the discs-large family of proteins are considered to be potential tumor suppressors. In the case of p55, this hypothesis has been strengthened by the finding that p55 interacts with a 30-kDa domain located in the N-terminal end of protein 4.1, a domain that is also present in the NF2 (neurofibromatosis 2) tumor suppressor gene (Marfatia *et al.,* 1994). Furthermore, the junctional plaque proteins, which include dlg-A, PSP95-SAP90, ZO-1, and ZO-2, are thought to be directly involved in the regulation of the organization of the cytoskeleton, cell adhesion, and cell motility and thus are considered to be potential tumor suppressors (Tsukita *et al.,* 1993).

The *DLG2* gene is located in a region showing LOH in about 25% of breast sporadic cancers (Nagai *et al.,* 1994; Cropp *et al.,* 1994; Futreal *et al.,* 1994). Until it was cloned, BRCA1, the breast and ovary cancer susceptibility gene located on 17q, was thought to be the target for these losses. However, analysis of 32 sporadic breast and 12 sporadic ovarian tumors displaying LOH for a marker located in an intron of BRCA1 failed to reveal any somatic mutations in BRCA1 (Futreal *et al.,* 1994). This has raised the possibility that a second gene located nearby might be involved in sporadic breast and ovarian tumors (Ponder, 1994; Vogelstein and Kinzler, 1994). Since *DLG2* was a good candidate, we investigated its involvement in sporadic breast cancer. Although a total of five different variants were found in four different exons when performing a mutation analysis, which demonstrates the reliability of the techniques used, no variations in migration of PCR fragments generated from tumor DNA alone were observed. Although it could be argued that mutations were missed, we believe that the combination of SSCP and heteroduplex analyses is sensitive enough to conclude that *DLG2* is not mutated in breast tumors. However, the involvement of *DLG2* in other cancers as a tumor suppressor warrants further attention.

ACKNOWLEDGMENTS

We thank P. Russel, P. Harrington, P. Holik, and J. Stevens for assistance and I. Mikaélian for helpful suggestions. This work was supported by the Cancer Research Campaign (S.A.G., A.D., and B.A.J.P.). S.M. is supported by an EC fellowship, S.A.S. is supported by an EMBO fellowship, and E.J.v.R. is the holder of the Lady Cade Memorial fellowship funded jointly by the CRC and National Cancer Association of South Africa. B.A.J.P. is a CRC Gibb fellow.

Note added in proof. Because a new human gene called hdlg (HGM symbol *DLG1*), which codes for a protein that belongs to the discs-large family, was published after this article was submitted (R. A. Lue, S. M. Marfatia, D. Branton, and A. H. Chishti (1994). Cloning and characterization of hdlg: The human homologue of the *Drosophila* discs large tumor suppressor binds to protein 4.1. *Proc. Natl. Acad. Sci. USA* **91:** 9818–9822), the gene that we report here has been called *DGL2*.

REFERENCES

Albertsen, H. M., Smith, S. A., Mazoyer, S., Fujimoto, E., Stevens, J., Williams, B., Rodriguez, P., Cropp, C. S., Slijepcevic, P., Carlson, M., Robertson, M., Bradley, P., Lawrence, E., Harrington, T., Mei Sheng, Z., Hoopes, R., Sternberg, N., Brothman, A., Callahan, R., Ponder, B. A. J., and White, R. (1994). A physical map and candidate genes in the *BRCA1* region on chromosome 17q12–21. *Nature Genet.* **7:** 472–479.

Bar-Sagi, D., Rotin, D., Batzer, A., Mandiyan, V., and Schlessinger, J. (1993). SH3 domains direct cellular localization of signaling molecules. *Cell* **74:** 83–91.

Bryant, P. J., Watson, K. L., Justice, R. W., and Woods, D. F. (1993). Tumor suppressor genes encoding proteins required for cell interactions and signal transduction in *Drosophila*. *Dev. Suppl.* 239–249.

Cho, K.-O., Hunt, C. A., and Kennedy, M. B. (1992). The rat brain postsynaptic density fraction contains a homolog of the Drosophila discs-large tumor suppressor protein. *Neuron* **9:** 929–942.

Cropp, C. S., Nevanlinna, H. A., Pyrhonen, S., Stenman, U.-H., Salmikangas, P., Albertsen, H., White, R., and Callahan, R. (1994). Evidence for involvement of *BRCA1* in sporadic breast carcinomas. *Cancer Res.* **54:** 2548–2551.

Duclos, F., Boeschert, U., Sirugo, G., Mandel, J. L., Hen, R., and Koenig, M. (1993). Gene in the region of the Friedreich ataxia locus encodes a putative transmembrane protein expressed in the nervous system. *Proc. Natl. Acad. Sci. USA* **90:** 109–113.

Futreal, P. A., Liu, Q., Shattuck-Eidens, D., Cochran, C., Harshman, K., Tavtigian, S., Bennett, L. M., Haugen-Strano, A., Swensen, J., Miki, Y., Eddington, K., McClure, M., Frye, C., Weaver-Feldhaus, J., Ding, W., Gholami, Z., Söderkvist, P., Terry, L., Jhanwar, S., Berchuck, A., Iglehart, J. D., Marks, J., Ballinger, D. G., Barrett, J. C., Skolnick, M., Kamb, A., and Wiseman, R. (1994). *BRCA1*

mutations in primary breast and ovarian carcinomas. *Science* **266:** 120–122.

Gorbalenya, A. E., and Koonin, E. V. (1990). Superfamily of UvrA-related NTP-binding proteins: Implications for rational classification of recombination/repair system. *J. Mol. Biol.* **213:** 583–591.

Gu, M. X., York, J. D., Warshawsky, I., and Majerus, P. W. (1991). Identification, cloning, and expression of a cytosolic megakaryocyte protein-tyrosine-phosphatase with sequence homology to cytoskeletal protein 4.1. *Proc. Natl. Acad. Sci. USA* **88:** 5867–5871.

Itoh, M., Nagafuchi, A., Yonemura, S., Kitani-Yasuda, T., Tsukita, S., and Tsukita, S. (1993). The 220-kD protein colocalizing with cadherins in non-epithelial cells is identical to ZO-1, a tight junction-associated protein in epithelial cells: cDNA cloning and immunoelectron microscopy. *J. Cell Biol.* **121:** 491–502.

Jacobs, I. J., Smith, S. A., Wiseman, R. W., Futreal, P. A., Harrington, T., Osborne, R. J., Leech, V., Molyneux, A., Berchuck, A., Ponder, B. A. J., and Bast, R. C. (1993). A deletion unit on chromosome 17q in epithelial ovarian tumors distal to the familial breast/ovarian cancer locus. *Cancer Res.* **53:** 1218–1221.

Jesaitis, L. A., and Goodenough, D. A. (1994). Molecular characterization and tissue distribution of ZO-2, a tight junction protein homologous to ZO-1 and the *Drosophila* discs-large tumor suppressor protein. *J. Cell Biol.* **124:** 949–961.

Kistner, U., Wenzel, B. M., Veh, R. W., Cases-Langhoff, C., Garner, A. M., Appeltauer, U., Voss, B., Gundelfinger, E. D., and Garner, C. C. (1993). SAP90, a rat presynaptic protein related to the product of the *Drosophila* tumor suppressor gene *dlg*-A. *J. Biol. Chem.* **268:** 4580–4583.

Koonin, E. V., Woods, D. F., and Bryant, P. J. (1992). dlg-R proteins: Modified guanylate kinases. *Nature Genet.* **2:** 256–257.

Marfatia, S. M., Lue, R. A., Branton, D., and Chishti, A. H. (1994). *In vitro* binding studies suggest a membrane-associated complex between erythroid p55, protein 4.1, and glycophorin C. *J. Biol. Chem.* **269:** 8631–8634.

Mayer, B. J., and Baltimore, D. (1993). Signaling through SH2 and SH3 domains. *Trends Cell Biol.* **3:** 8–13.

Metzenberg, A. B., and Gitschier, J. (1992). The gene encoding the palmitoylated erythrocyte membrane protein, p55, originates at the CpG island 3′ to the factor VIII gene. *Hum. Mol. Genet.* **1:** 97–101.

Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P. A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L. M., Ding, W., Bell, R., Rosenthal, J., Hussey, C., Tran, T., McClure, M., Frye, C., Hattier, T., Phelps, R., Haugen-Strano, A., Katcher, H., Yakumo, K., Gholami, Z., Shaffer, D., Stone, S., Bayer, S., Wray, C., Bogden, R., Dayananth, P., Ward, J., Tonin, P., Narod, S., Bristow, P. K., Norris, F. H., Helvering, L., Morrison, P., Rosteck, P., Lai, M., Barrett, J. C., Lewis, C., Neuhausen, S., Cannon-Albright, L., Goldgar, D., Wiseman, R., Kamb, A., and Skolnick, M. H. (1994). A strong candidate for the breast and ovarian cancer susceptibility gene *BRCA1*. *Science* **266:** 66–71.

Musacchio, A., Gibson, T., Lehto, V.-P., and Saraste, M. (1992). SH3—An abundant protein domain in search of a function. *FEBS Lett.* **307:** 55–61.

Nagai, M. A., Yamamoto, L., Salaorni, S., Pacheco, M. M., Brentani, M. M., Barbosa, E. M., Brentani, R. R., Mazoyer, S., Smith, S. A., Ponder, B. A. J., and Mulligan, L. M. (1994). Detailed deletion mapping of chromosome segment 17q12–21 in sporadic breast tumours. *Genes Chrom. Cancer* **11:** 58–62.

Ponder, B. A. J. (1994). Searches begin and end. *Nature* **371:** 279.

Ruff, P., Speicher, D. W., and Husain-Chishti, A. (1991). Molecular identification of a major palmitoylated erythrocyte membrane protein containing the *src* homology 3 motif. *Proc. Natl. Acad. Sci. USA* **88:** 6595–6599.

Saito, H., Inazawa, J., Saito, S., Kasumi, F., Koi, S., Sagae, S., Kudo, R., Saito, J., Noda, K., and Nakamura, Y. (1993). Detailed deletion mapping of chromosome 17q in ovarian and breast cancer: 2 cM region on 17q21.3 often and commonly deleted in tumors. *Cancer Res.* **53:** 3382–3385.

Stehle, T., and Schulz, G. E. (1990). Three-dimensional structure of the complex of guanylate with its substrate GMP. *J. Mol. Biol.* **211:** 249–254.

Tsukita, S., Itoh, M., Nagafuchi, A., Yonemura, S., and Tsukita, S. (1993). Submembranous junctional plaque proteins include potential tumor suppressor molecules. *J. Cell Biol.* **123:** 1049–1053.

Vogelstein, B., and Kinzler, K. W. (1994). Has the breast cancer gene been found? *Cell* **79:** 1–3.

Willott, E., Balda, M. S., Fanning, A. S., Jameson, B., Van Itallie, C., and Anderson, J. M. (1993). The tight junction protein ZO-1 is homologous to the *Drosophila* discs-large tumor suppressor protein of septate junctions. *Proc. Natl. Acad. Sci. USA* **90:** 7834–7838.

Woods, D. F., and Bryant, P. J. (1991). The discs-large tumor suppressor gene of Drosophila encodes a guanylate kinase homolog localized at septate junctions. *Cell* **66:** 451–464.

GENOMICS 31, 145–150 (1996)
ARTICLE NO. 0025

Isolation of a Gene (DLG3) Encoding a Second Member of the Discs-Large Family on Chromosome 17q12–q21

SIMON A. SMITH,[1] PILAR HOLIK, JEFF STEVENS, SYLVIE MAZOYER,* ROBERTA MELIS, BRIANA WILLIAMS, RAY WHITE, AND HANS ALBERTSEN

*Huntsman Cancer Institute, University of Utah, Salt Lake City, Utah 84112; and *CRC Human Cancer Genetics Research Group, University of Cambridge, Addenbrooke's Hospital, Cambridge CB2 2QQ, United Kingdom*

Received June 8, 1995; accepted October 30, 1995

The discs-large family is a collection of proteins that have a common structural organization and are thought to be involved in signal transduction and mediating protein–protein interactions at the cytoplasmic surface of the cell membrane. The defining member of this group of proteins is the gene product of the Drosophila lethal (1) discs large (dlg) 1 locus, which was originally identified by the analysis of recessive lethal mutants. Germline mutations in dlg result in loss of apical–basolateral polarity, disruption of normal cell–cell adhesion, and neoplastic overgrowth of the imaginal disc epithelium. We have isolated and characterized a novel human gene, DLG3, that encodes a new member of the discs-large family of proteins. The putative DLG3 gene product has a molecular weight of 66 kDa and contains a discs-large homologous region, a src oncogene homology motif 3, and a domain with homology to guanylate kinase. The DLG3 gene is located on chromosome 17, in the same segment, 17q12–q21, as the related gene, DLG2. The products of the DLG2 and DLG3 genes show 36% identity and 58% similarity to each other, and both show nearly 60% sequence similarity to p55, an erythroid phosphoprotein that is a component of the red cell membrane. We suggest that p55, DLG2, and DLG3 are closely related members of a gene family, whose protein products have a common structural organization and probably a similar function.

INTRODUCTION

The discs-large family is an expanding group of proteins that each contain three distinct structural domains: an N-terminal segment comprising one or more discs-large homologous regions (DHRs); a src oncogene homology motif 3 (SH3); and a C-terminal domain with homology to guanylate kinase (GK). The founding member of this group of proteins is the gene product of the Drosophila lethal (1) discs large (dlg) 1 locus, which was originally identified through the characterization of recessive lethal lesions that lead to overproliferation of the imaginal disc epithelium and death of the fly in the larval stage (Stewart *et al.*, 1972; Woods and Bryant, 1989, 1991). At all stages of development, the dlg protein appears to be a component of the septate junctions (Woods and Bryant, 1991), which are found in the apical–lateral membrane of epithelial cells and are thought to be functionally equivalent to vertebrate tight junctions (Noirot-Timothee and Noirot, 1980). Germline mutations in dlg result in loss of apical–basolateral polarity, disruption of normal cell–cell adhesion, and neoplastic overgrowth of the imaginal disc epithelium (Stewart *et al.*, 1972; Woods and Bryant, 1989), suggesting that the function of the wildtype dlg protein is concerned with maintaining normal epithelial structure and possibly with regulating cell division and differentiation (Woods and Bryant, 1991). Other members of the discs-large family include ZO-1 and ZO-2, which are mammalian tight junction proteins (Willott *et al.*, 1993; Jesaitis and Goodenough, 1994); SAP90/PSD-95, a rat synapse-associated protein (Cho *et al.*, 1992; Kistner *et al.*, 1993); p55, an erythroid phosphoprotein that is a component of the red cell membrane (Ruff *et al.*, 1991); and hdlg, a recently discovered human homolog of Drosophila dlg (Lue *et al.*, 1994).

The DHR motif is approximately 90 amino acids long and appears to be involved in forming protein–protein interactions. *In vitro* binding studies have demonstrated that hdlg binds to protein 4.1 (Lue *et al.*, 1994), the defining member of a group of proteins that includes talin (Rees *et al.*, 1990) and ezrin (Gould *et al.*, 1989), which are thought to couple the cytoskeleton to the cell membrane (Marfatia *et al.*, 1994). The binding of protein 4.1 to hdlg was mapped to two sites on hdlg: a domain containing all three DHR segments and an insertion domain (I3) found between the SH3 and GK

The DNA sequence presented in this article has been deposited with the EMBL/GenBank Data Libraries under Accession No. U37707.

[1] To whom correspondence should be addressed at Huntsman Cancer Institute, University of Utah, Building 533, Suite 7410, Salt Lake City, Utah 84112. Telephone: (801) 585-6224. Fax: (801) 585-3833.

0888-7543/96 $12.00

domains in a subset of hdlg isoforms (Lue *et al.*, 1994). In human erythrocytes, spectrin binds to the C-terminal end of protein 4.1, while at the N-terminus, protein 4.1 forms a ternary complex with p55 and glycophorin C (a transmembrane protein) that links the spectrin network to the cell membrane (Marfatia *et al.*, 1994). However, it is not clear what the function of p55 is in this process, and the binding sites in p55 have not been precisely mapped. The SH3 domain comprises about 60 amino acids and was originally identified in the noncatalytic region of the src family of nonreceptor protein tyrosine kinases (Pawson, 1988). SH3 domains have now been found in a wide variety of proteins, including signal-transducing proteins and membrane-associated cytoskeletal elements, and are thought to be involved in mediating protein–protein interactions by binding to specific target sequences (Koch *et al.*, 1991; Mayer and Eck, 1995). The GK domain is homologous to the complete amino acid sequence of yeast guanylate kinase (Berger *et al.*, 1989), an enzyme that catalyzes the transfer of phosphate from ATP to GMP, forming GDP, although at the moment, it is unclear whether any of the dlg-related proteins have guanylate kinase activity.

Recently, we reported the isolation and characterization of a gene, DLG2, that encodes another novel member of the discs-large family of proteins (Mazoyer *et al.*, 1995). The predicted DLG2 gene product contains a single DHR motif and shows greatest homology to p55. Here, we describe the isolation of a second novel human gene, DLG3, located close to DLG2 on chromosome segment 17q12–q21, which is predicted to encode a protein that is nearly 60% homologous to the p55 and DLG2 gene products. We suggest that p55, DLG2, and DLG3 are three members of a family of genes that encode proteins of similar size and structural organization that are probably involved in coupling the cytoskeleton to the cell membrane.

MATERIALS AND METHODS

DLG3 cDNA isolation and DNA sequencing. Genomic DNA clones that form part of a physical contig surrounding the BRCA1 locus (Albertsen *et al.*, 1994) were used to screen a fetal brain cDNA library to identify transcribed sequences in the interval 17q12–q21. Over 100 cDNAs were isolated, representing transcripts of known and unknown genes, several of which have been reported elsewhere (Albertsen *et al.*, 1994; Mazoyer *et al.*, 1995; Smith *et al.*, 1995). A clone, 40F1, encoded part of the transcript of another novel gene, which has not been previously described, and was used to isolate further overlapping cDNAs. Inspection of the combined DNA sequence from several clones did not reveal a candidate initiation codon, suggesting that the extreme 5′ end of the cDNA remained to be isolated. RACE (Frohman *et al.*, 1988) experiments extended the sequence in the 5′ direction for a further 100 bp and revealed a candidate ATG located close to the 5′ end of the sequence. Careful rescreening of cDNA libraries failed to identify any additional clones, and thus we sought to isolate further sequence information from genomic DNA directly. Analysis of subclones generated from the P1 phage, 124D3, yielded a further 500 bp of 5′ sequence that confirmed the presence of the in-frame ATG and identified stop codons in all three frames upstream.

Analysis of DLG3 gene expression. A blot of RNAs extracted from different tissues was purchased from Clontech and hybridized as described by the manufacturer. A 1.8-kb *Eco*RI restriction DNA fragment excised from the 40F1 cDNA was used as the probe.

Database searches and sequence comparisons. To identify homologies with nucleic acid and protein sequences in the GenBank and EMBL databases, we used the BLAST algorithm (Altschul *et al.*, 1990). Sequences were aligned and compared using the Wisconsin Sequence Analysis Package Version 8.0.

RESULTS AND DISCUSSION

A novel cDNA clone (40F1) derived from human chromosome segment 17q12–q21 was isolated from a fetal brain cDNA library and used to identify further overlapping clones. The combined DNA sequence from several clones revealed a single long open reading frame (ORF) that extended from the most 5′ end of the sequence to a stop codon located over 1700 bp downstream. The cDNA contained 900 bp of 3′-nontranslated sequence and terminated with a poly(A) tail. Since the DNA sequence did not contain a candidate initiation codon, the 5′ end of the cDNA was pursued using the RACE technique, which extended the 5′ sequence for a further 100 bp and revealed a probable start codon. The completed ORF extended over 1755 bp (Fig. 1), encoding a protein of 585 amino acids with a predicted molecular weight of 66 kDa. The expected gene product represents a new member of the discs-large family of proteins (see below), and the newly isolated gene has been designated DLG3.

The molecular characterization of DLG3 included an analysis of the genomic structure of the gene. Intron–exon boundaries were identified as points of discontinuity between the cDNA sequence and portions of genomic sequence. Fragments of human DNA containing potential boundaries were generated either by PCR using primer pairs located within adjacent coding exons or by subcloning fragments from P1 phage. The coding sequence was found to be distributed between 15 exons, the boundaries of which are indicated in Fig. 1. The 9 3′-most exons were found to span a genomic interval of about 16 kb (based on the combined sizes of genomic clones and PCR products); the 3 5′-most exons were contained within a 1200-bp region; and the genomic extent of exons 4, 5, and 6 was not clearly defined.

Comparison of DLG3 with sequences present in the EMBL and GenBank databases revealed significant homology, at the amino acid level, to several members of the discs-large family of proteins. No homologies to any known genes were observed at the nucleotide level, although identity to three anonymous genomic DNA sequences (T27219, T27220, and T27221) was noted (see below). The greatest protein homology observed was to human erythrocyte p55, with which it shared 33% identity and 56% similarity. Alignment of the predicted amino acid sequence of the DLG3 protein with other members of the discs-large family using the PILE-UP program revealed that DLG3 contains a single DHR, an SH3 motif, and a guanylate kinase domain (Fig. 2). The DHR in DLG3 is, however, not well conserved: it shares only 30% identity to the third repeat in dlg, and

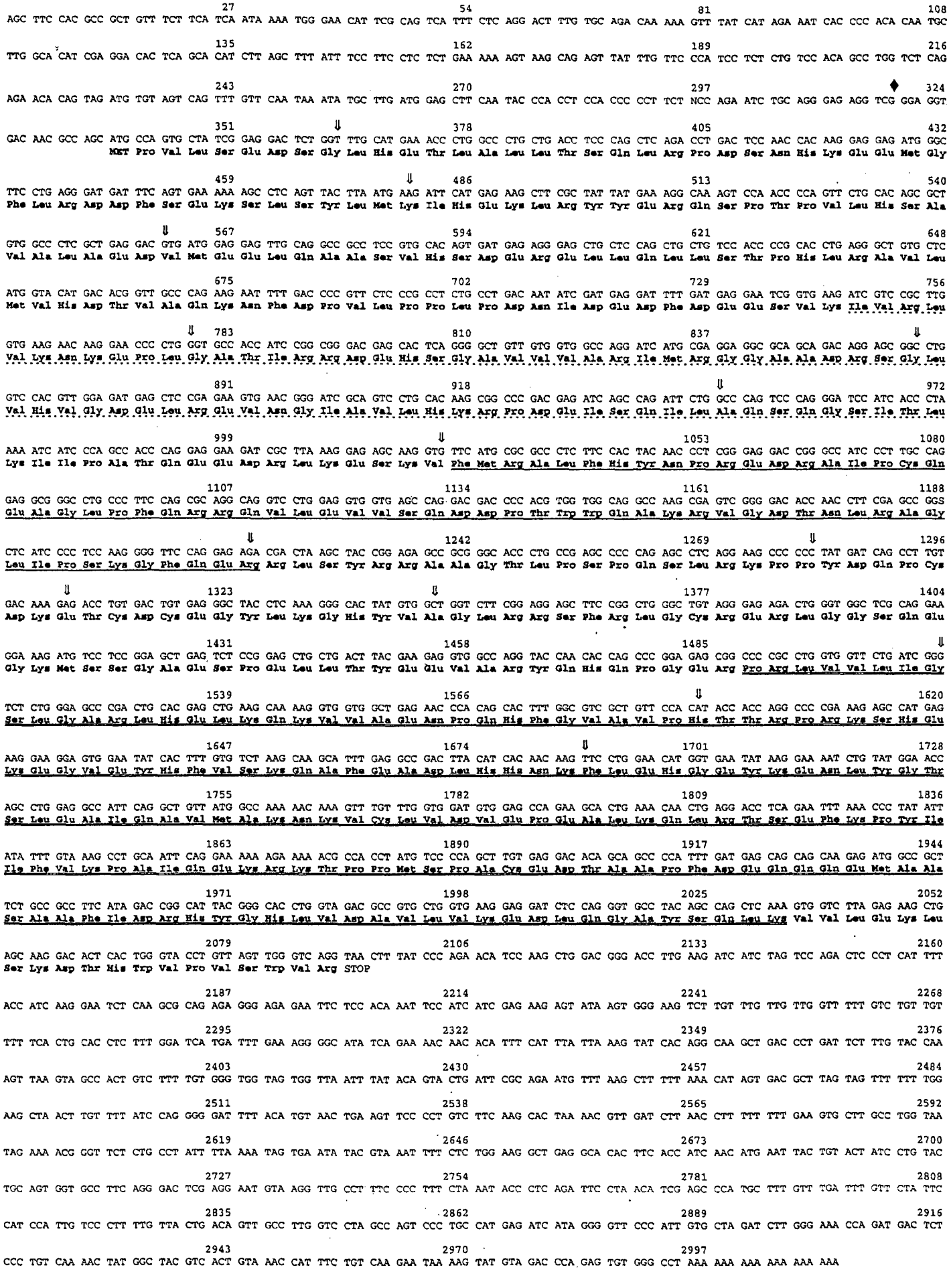


FIG. 1. Nucleotide sequence and predicted protein translation of the DLG3 gene. The dashed line indicates the position of the DHR, the solid line indicates the SH3 domain, and the double underline indicates the GK domain. The vertical arrows in the coding sequence indicate the positions of exonic boundaries, and the black diamond identifies the point 5′ from where the sequence was derived from genomic DNA.

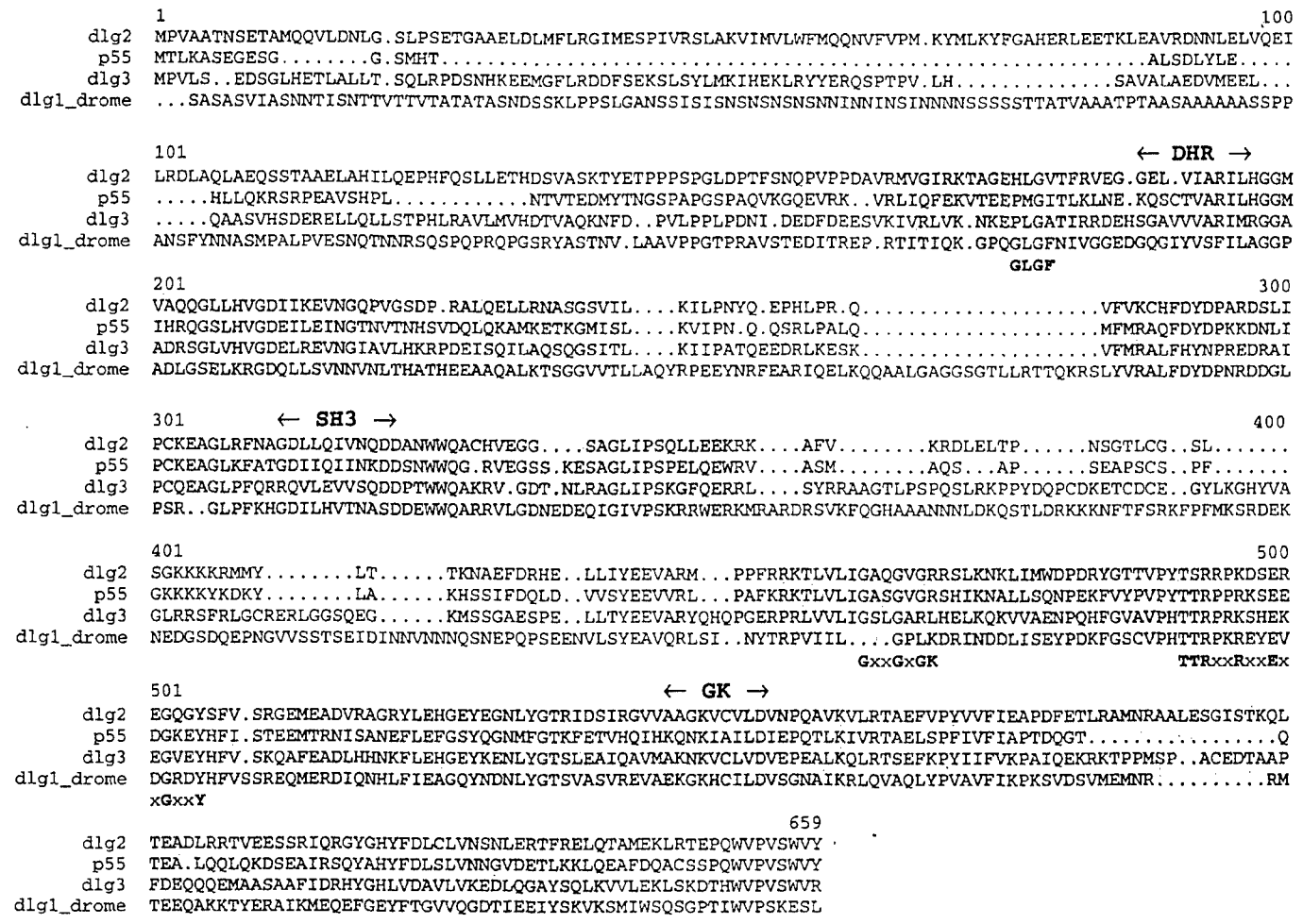


FIG. 2. Comparison of the predicted amino acid sequences of the p55, DLG2, DLG3, and Drosophila dlg gene products. The putative amino acid sequences are complete with the exception of dlg, for which the N-terminal 325 residues were deleted for clarity. The sequences were compared using the PILEUP program (Wisconsin Sequence Analysis Package Version 8.0) with a gap weight of 3.6 and a length weight of 0.30. The shading indicates the locations of the DHR, SH3, and GK domains. The sequence "GLGF" indicates the position of this motif in Drosophila dlg (see text); GxxGxGK and TTRxxRxxExxGxxY are consensus sequences (where x is any amino acid) marking the positions, in the GK domain, of the "anion hole" and the "GMP-binding motif," respectively.

the sequence motif, GLGF (Cho *et al.,* 1992), which is conserved in dlg, hdlg, and PSD-95, is absent in DLG3. The N-terminal region of the DLG3 protein most resembles p55; both proteins contain only one DHR (whereas the other members contain three), and the DHR in both proteins is poorly conserved compared with dlg. In contrast, the SH3 motif appears to be highly conserved between all members of the discs-large family, including the DLG3 gene product. Two regions of the guanylate kinase domain deserve close inspection, the "anion hole" and the "GMP-binding site" (Koonin *et al.,* 1992; Willott *et al.,* 1993). The anion hole comprises the signature GxxGxGKST, which is well conserved in p55, but contains a deletion in dlg, hdlg, and PSD-95. In DLG3, the spacing is correct, but the GKST residues are not retained, and thus it is unclear whether the putative DLG3 protein is a kinase. The GMP-binding site contains the consensus sequence, TTRxxRxxExxGxxY, which is conserved in all dlg-related proteins with the exception of ZO-1 and ZO-2 and is present in DLG3, although the second arginine residue is replaced by a lysine, as it is in p55. Thus, the inclusion of DLG3 as a new member of the discs-large family seems reasonable based upon the functional domains that it contains, which are common to all discs-large proteins.

Analysis of DLG3 gene expression revealed that the transcript is present in all tissues tested, except for peripheral blood leukocytes (Fig. 3). At least four tissue-specific transcripts were identified, ranging from about 2 to 4.5 kb. Thymus contained approximately equal proportions of two transcripts of 3.5 and 4.5 kb, while spleen, prostate, ovary, small intestine, and colon contained an additional (though frequently poorly expressed) 2.5-kb transcript. In testis only, a 2-kb mRNA was identified (in addition to the three other transcripts) and was the most abundantly expressed transcript in this tissue. Since the largest transcript identified was about 4.5 kb, and the length of the cDNA from the presumed ATG start codon to the poly(A) tail is only 2650 bp, it is not certain that the complete coding sequence has been identified. RACE experiments and rescreening of libraries did not extend the cDNA sequence further in the 5′ direction, and thus the length

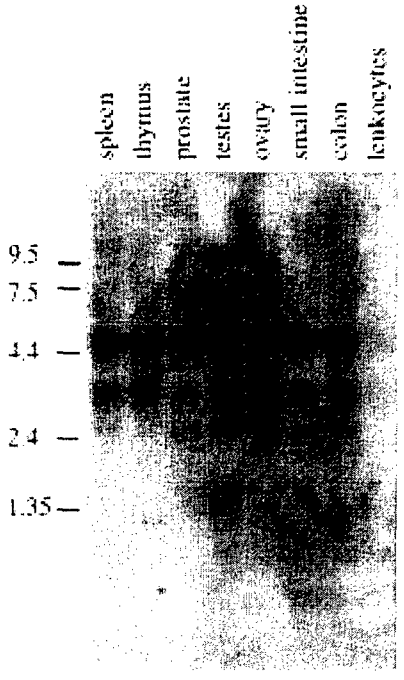


FIG. 3. Analysis of the expression of the DLG3 gene in different tissues. A blot of RNAs extracted from various tissues was hybridized with a radiolabeled DNA fragment from the 40F1 cDNA that contained most of the coding sequence of DLG3. No bands were identified in the lane containing RNA from peripheral leukocytes; hybridization of the same blot with an actin probe indicated that each lane was loaded with an equivalent amount of mRNA.

of the 5′-untranslated region (UTR) is not known. In any case, it is unlikely that the 5′ UTR would account for the discrepancy between the size of the transcript and the length of the coding sequence. Other possibilities include alternative 3′ polyadenylation sites that increase the length of the 3′ UTR, or the ATG codon that we identified as the translational start of the gene may in fact be downstream of the true initiator, although this is unlikely (see below). Another explanation is that other coding exons that are not expressed in fetal brain exist, and thus were spliced out of the cDNA clones that we had isolated. This explanation is all the more plausible given the observation that most tissues tested contain several different DLG3 transcripts, although it is not known which variants are expressed in fetal brain.

Recently, we reported another gene located on chromosome 17, DLG2, that also encodes a protein belonging to the discs-large family (EMBL Accession No. X82895; Mazoyer *et al.,* 1995). The DLG2 and DLG3 genes have both been mapped to the 280-kb YAC, 853B3, whose location on chromosome 17q has been confirmed by fluorescence *in situ* hybridization (FISH) (Albertsen *et al.,* 1994). P1 clones containing the DLG2 and DLG3 genes have also been isolated and mapped to 17q by FISH, although attempts at isolating clones that contain all or part of both genes have been unsuccessful. The colocalization of DLG2 and DLG3 on 853B3 certainly suggests that they are located close together on the chromosome, although the precise distance between them and the orientation of the genes with respect to each other is not known. Finally, the observation that three genomic sequences present in the EMBL and GenBank databases (EMBL Accession Nos. T27219, T27220, and T27221), which are also derived from 17q12–q21, share identity with small regions of the DLG3 cDNA sequence supports the localization of the DLG3 gene on chromosome 17.

Comparison of the predicted DLG3 protein with that of DLG2 showed that they contain similar structural domains: both proteins contain a single DHR, an SH3 motif, and a guanylate kinase domain containing a highly conserved GMP-binding site. Overall, the amino acid sequences of the two proteins share 36% identity and 58% similarity, which is close to that observed for both DLG2 and DLG3 compared individually with p55. The relatedness of DLG2, DLG3, and p55 suggests a common functionality, possibly as "linker proteins" important in coupling the cytoskeleton to the cell membrane.

Another feature of the DLG2, DLG3, and hdlg proteins (though not p55) is that they are each predicted to start with the same three amino acids, MPV. While the function of the MPV signature is not known and the sequence match is unlikely to be coincidental, the presence of this motif at the start of the DLG3 protein suggests that the translational start site of the DLG3 gene has been correctly identified. The presumed ATG start codon also occurs within the consensus Kozak environment, GCC(A/G)CCATGG, which is found at the translational start site of most vertebrate genes (Kozak, 1991). At the C-terminus, the p55 and DLG2 proteins end with the sequence WVPVSWVY, while the DLG3 protein terminates with a near-identical sequence in which the tyrosine residue is replaced by an arginine. Again, the function of this C-terminal motif is not known, but its presence in the predicted DLG3 protein strongly indicates that the extreme 3′ end of the coding sequence has been correctly obtained.

Analysis of the nucleotide sequences of the p55, DLG2, and DLG3 genes using the BESTFIT program (Wisconsin Sequence Analysis Package Version 8.0), which introduces gaps into the sequences to obtain optimal alignment, reveals that they show approximately 60% identity to one another. The sequence homology among p55, DLG2, and DLG3 suggests that they belong to a "gene family," although it is not clear whether other family members exist and how the genes evolved. However, the close physical proximity of DLG2 and DLG3 (p55 is located on the X chromosome (Metzenberg and Gitschier, 1992)) suggests that the two genes evolved by a duplication event.

Recent experiments indicate that the p55 gene product is involved in coupling the cytoskeleton to the red cell membrane by forming a ternary complex with protein 4.1 and glycophorin C (Marfatia *et al.,* 1994). Protein 4.1 also binds to spectrin, a key component of the intracellular protein network that underlies the erythrocyte membrane. It is possible that the DLG2 and DLG3 gene products have a function similar to p55 in

helping to maintain the integrity of the cell membrane and its links with associated cytoplasmic proteins.

ACKNOWLEDGMENTS

We thank Margaret Robertson and Elizabeth Lawrence for DNA sequencing and Ed Meenen for synthesizing oligonucleotides. S.A.S. is supported by an EMBO fellowship. This work was supported in part by Grant DAMD17-94-J-4129 from the Department of Defence to R.W.

REFERENCES

Albertsen, H. M., Smith, S. A., Mazoyer, S., Fujimoto, E., Stevens, J., Williams, B., Rodriguez, P., Cropp, C. S., Slijepcevic, P., Carlson, M., Robertson, M., Bradley, P., Lawrence, E., Harrington, T., Sheng, Z. M., Hoopes, R., Sternberg, N., Brothman, A., Callahan, R., Ponder, B. A. J., and White, R. (1994). A physical map and candidate genes in the BRCA1 region on chromosome 17q12–21. *Nature Genet.* **7:** 472–479.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

Berger, A., Schiltz, E., and Schulz, G. E. (1989). Guanylate kinase from Saccharomyces cerevisiae. Isolation and characterisation, crystallization and preliminary x-ray analysis, amino acid sequence and comparison with adenylate kinases. *Eur. J. Biochem.* **184:** 433–443.

Cho, K.-O., Hunt, C. A., and Kennedy, M. B. (1992). The rat brain postsynaptic density fraction contains a homolog of the Drosophila discs-large tumour suppressor protein. *Neuron* **9:** 929–942.

Frohman, M. A., Dush, M. K., and Martin, G. R. (1988). Rapid production of full length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA* **85:** 8998–9002.

Gould, K. L., Bretscher, A., Esch, F. S., and Hunter, T. (1989). cDNA cloning and sequencing of the protein-tyrosine kinase substrate, ezrin, reveals homology to band 4.1. *EMBO J.* **8:** 4133–4142.

Jesaitis, L. A., and Goodenough, D. A. (1994). Molecular characterisation and tissue distribution of ZO-2, a tight junction protein homologous to ZO-1 and the Drosophila discs-large tumour suppressor protein. *J. Cell Biol.* **124:** 949–961.

Kistner, U., Wenzel, B. M., Veh, R. W., Cases-Langhoff, C., Garner, A. M., Appeltauer, U., Voss, B., Gundelfinger, E. D., and Garner, C. C. (1993). SAP90, a rat presynaptic protein related to the product of the Drosophila tumour suppressor gene dlg-A. *J. Biol. Chem.* **268:** 4580–4583.

Koch, C. A., Anderson, D., Moran, M. F., Ellis, C., and Pawson, T. (1991). SH2 and SH3 domains: Elements that control interactions of cytoplasmic signaling proteins. *Science* **252:** 668–674.

Koonin, E. V., Woods, D. F., and Bryant, P. J. (1992). dlg-R proteins: Modified guanylate kinases. *Nature Genet.* **2:** 256–257.

Kozak, M. (1991). An analysis of vertebrate mRNA sequences: Intimations of translational control. *J. Cell Biol.* **115:** 887–903.

Lue, R., Marfatia, S. M., Branton, D., and Chishti, A. H. (1994). Cloning and characterisation of hdlg: The human homolog of the Drosophila discs-large tumour suppressor binds to protein 4.1. *Proc. Natl. Acad. Sci. USA* **91:** 9818–9822.

Marfatia, S. M., Lue, R. A., Branton, D., and Chishti, A. H. (1994). In vitro binding studies suggest a membrane-associated complex between erythroid p55, protein 4.1 and glycophorin C. *J. Biol. Chem.* **269:** 8631–8634.

Mayer, B. J., and Eck, M. J. (1995). SH3 domains. Minding your p's and q's. *Curr. Biol.* **5:** 364–367.

Mazoyer, S. M., Gayther, S. A., Nagai, M. A., Smith, S. A., Dunning, A., van Rensburg, E. J., Albertsen, H., White, R., and Ponder, B. A. J. (1995). A gene (*DLG2*) located at 17q12–q21 encodes a new homologue of the Drosophila tumor suppressor dlg-A. *Genomics* **28:** 25–31.

Metzenberg, A. B., and Gitschier, J. (1992). The gene encoding the palmitoylated erythrocyte membrane protein, p55, originates at the CpG island 3′ to the factor VIII gene. *Hum. Mol. Genet.* **1:** 97–101.

Noirot-Timothee, C., and Noirot, C. (1980). Septate and scalariform junctions in arthropods. *Int. Rev. Cytol.* **63:** 97–140.

Pawson, T. (1988). Non-catalytic domains of cytoplasmic protein-tyrosine kinases: Regulatory elements in signal transduction. *Oncogene* **3:** 491–495.

Rees, D. J. G., Ades, S. E., Singer, S. J., and Hines, R. O. (1990). Sequence and domain structure of talin. *Nature (London)* **347:** 685–689.

Ruff, P., Speicher, D. W., and Chishti, A. H. (1991). Molecular identification of a major palmitoylated erythrocyte membrane protein containing the src homology 3 motif. *Proc. Natl. Acad. Sci. USA* **88:** 6595–6599.

Smith, S. A., Holik, P., Stevens, J., Melis, R., White, R., and Albertsen, H. (1995). Isolation and mapping of a gene encoding a novel human ADP-ribosylation factor on chromosome 17q12–q21. *Genomics* **28:** 113–115.

Stewart, M., Murphy, C., and Fristrom, J. (1972). The recovery and preliminary characterisation of X chromosome mutants affecting imaginal discs of *Drosophila melanogaster*. *Dev. Biol.* **27:** 71–83.

Willott, E., Balda, M. S., Fanning, A. S., Jameson, B., Van Itallie, C., and Anderson, J. M. (1993). The tight junction protein ZO-1 is homologous to the Drosophila discs-large tumour suppressor proteins of septate junctions. *Proc. Natl. Acad. Sci. USA* **90:** 7834–7838.

Woods, D. F., and Bryant, P. J. (1989). Molecular cloning of the lethal (1) discs-large (1) oncogene of Drosophila. *Dev. Biol.* **134:** 222–235.

Woods, D. F., and Bryant, P. J. (1991). The discs-large tumour suppressor gene of Drosophila encodes a guanylate kinase homolog localised at septate junctions. *Cell* **66:** 451–464.

SHORT COMMUNICATION

Isolation and Mapping of a Gene Encoding a Novel Human ADP-Ribosylation Factor on Chromosome 17q12–q21

SIMON A. SMITH,[1] PILAR R. HOLIK, JEFF STEVENS, ROBERTA MELIS, RAY WHITE, AND HANS ALBERTSEN

Huntsman Cancer Institute, University of Utah, Building 533, Suite 7410, Salt Lake City, Utah

Received December 12, 1994; accepted April 19, 1995

A gene encoding a low-molecular-weight GTP-binding protein was isolated from a retinal cDNA library and mapped to human chromosome 17q12–q21. Comparison of the predicted protein with the protein databases revealed striking homology to the family of ADP-ribosylation factors (ARFs), which are thought to be involved in membrane trafficking and protein secretion. The greatest homology observed was with the rat ARF-like 4 protein (ARL4), with which it shared 58% identity, while the more highly conserved human ARF1 and ARF3 proteins each shared 46% identity. Inspection of the predicted new protein showed that it contained each of the six conserved motifs that are required for guanine nucleotide binding and hydrolysis, and thus it is probably a novel ARF isoform. We have designated the new protein and its corresponding gene ARF4L.

ADP-ribosylation factors (ARFs) are a subfamily of low-molecular-weight GTP-binding proteins that are thought to be involved in membrane trafficking and protein secretion (2, 4, 6). Each protein has a predicted molecular weight of approximately 20 kDa and contains six distinct domains that are required for nucleotide binding and hydrolysis (7). Three domains (PM1–3) are required for phosphate/magnesium binding, while the other three domains (G1–3) are required for guanine nucleotide binding. The genes for at least six human ARF isoforms have been identified (3); comparison of the deduced amino acid sequences indicates a high degree of sequence conservation. Mouse homologs of the human genes have also been isolated, and two rat genes, encoding ARF-like (ARL) proteins ARL1 and ARL4, have been identified (5). The predicted ARL1 and ARL4 proteins also contain the six conserved motifs required for GTP-binding and hydrolysis, but share only 55 and 41% identity to ARF1, respectively. Here, we report the cloning of a novel human ARF gene that is most homologous to rat ARL4, but that probably represents a distinct ARL isoform that maps to chromosome 17q12–q21.

A retinal cDNA library (Stratagene 937202) was screened by hybridization with P1 phage clones located on 17q12–q21. One of several positive clones that mapped back to chromosome 17q, 16RB1, contained an insert of 1376 bp and had a short poly(A) tail at one end. Sequence analysis revealed a single long open reading frame of 603 bp that was predicted to encode a protein that has a molecular weight of 22.3 kDa (Fig. 1). Comparison of the nucleotide sequence of the 16RB1 clone with the EMBL and NCBI databases using the BLAST algorithm revealed identity to the partial sequence of a human cDNA clone, c-29a01 (Accession Nos. Z40585 and Z44776), while the predicted amino acid sequence revealed striking homology to ARF and ARF-like proteins from several different species. The greatest protein homology observed was with rat ARL4, which showed 58% identity and 78% similarity (Fig. 2). In contrast, the highly conserved ARF proteins were less homologous to the predicted new protein: human ARF1 and ARF3 showed 46% identity to the new isoform. Since the predicted new protein meets the criteria for belonging to the ARF family of GTP-binding proteins (see below) and the protein appears to represent a previously unknown isoform, it has been designated the name ARF4L.

Inspection of the predicted amino acid sequence of the ARF4L protein reveals that it contains each of the motifs required for guanine nucleotide binding and hydrolysis (Fig. 2). While the motifs themselves match the consensus very closely (with the exception of G3, which appears most like the human ARL2 isoform), the spacing between the PM2 and the PM3 domains in ARF4L is especially noteworthy, since it is similar to that in rat ARL4: both of these proteins contain an extra five amino acids that are absent from the other ARF and ARF-like proteins. In common with the other members of the ARF subfamily, ARF4L contains a glycine at position 2 and does not contain any cysteine residues in the C-terminus of the protein, which are common in other GTP-binding proteins. Thus, the inclusion of the ARF4L protein as a member of the ARF subfamily seems reasonable based upon its alignment with known members of this family.

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession No. L38490.

[1] To whom correspondence should be addressed. Telephone: (801) 585-6224. Fax: (801) 585-3833.

GENOMICS **28,** 113–115 (1995)
0888-7543/95 $12.00

The precise functions of the ARF and ARF-like proteins are unknown. Several studies have indicated that some of the ARF genes show tissue-specific expression. For instance, ARF5 is expressed predominantly in brain, and ARF4 is expressed in adipose tissue but not

```
                                    28                                  55
gca cga ggg tgt ctg cgg ggg tct cgc ggg gcg gct gcg gtg ttt cac cgg gga
                                    82                                 109
aag gct cga gga gag cgc ggc tca cga gag ata acc cag ctg tgc tcc ctg gaa
                                   136                                 163
cct tca att tca agg cct ccc tgc ctc tac tag gcg cct tag ctc act ATG GGG
                                                                   MET Gly
                                   190                                 217
AAC CAC TTG ACT GAG ATG GCG CCC ACT GCC TCC TCC TTC TTG CCC CAC TTC CAA
Asn His Leu Thr Glu Met Ala Pro Thr Ala Ser Ser Phe Leu Pro His Phe Gln
                                   244                                 271
GCC CTG CAT GTC GTG GTC ATT GGG CTG GAC TCT GCT GGA AAG ACC TCC CTC CTT
Ala Leu His Val Val Val Ile Gly Leu Asp Ser Ala Gly Lys Thr Ser Leu Leu
                                   298                                 325
TAC CGS CTC AAG TTC AAG GAG TTT GTC CAG AGT GTC CCC ACC AAA GGC TTC AAC
Tyr Arg Leu Lys Phe Lys Glu Phe Val Gln Ser Val Pro Thr Lys Gly Phe Asn
                                   352                                 379
ACC GAG AAG ATC CGG GTG CCC CTC GGG GGA TCG CGT GGC ATC ACC TTC CAA GTG
Thr Glu Lys Ile Arg Val Pro Leu Gly Gly Ser Arg Gly Ile Thr Phe Gln Val
                                   406                                 433
TGG GAC GTC GGG GGG CAG GAG AAG CTG CGA CCA CTG TGG CGC TCT TAT AAC CGC
Trp Asp Val Gly Gly Gln Glu Lys Leu Arg Pro Leu Trp Arg Ser Tyr Asn Arg
                                   460                                 487
CGG ACA GAC GGT CTA GTG TTT GTG GTG GAC GCT GCG GAG GCT GAG CGG CTT GAG
Arg Thr Asp Gly Leu Val Phe Val Val Asp Ala Ala Glu Ala Glu Arg Leu Glu
                                   514                                 541
GAA GCM AAG GTG GAG TTG CAC CGA ATC AGC CGG GCC TCG GAC AAC CAG GGC GTG
Glu Ala Lys Val Glu Leu His Arg Ile Ser Arg Ala Ser Asp Asn Gln Gly Val
                                   568                 -------------
CCA GTS CTG GTG CTG GCC AAC AAG CAG GAC CAG CCC GGG GCA CTG AGC GCT GCT
Pro Val Leu Val Leu Ala Asn Lys Gln Asp Gln Pro Gly Ala Leu Ser Ala Ala
-S2AF-------->                      622                                 649
GAG GTG GAG AAG AGG CTG GCA GTC CGA GAG CTA GCA GCC GCC ACT CTC ACT CAT
Glu Val Glu Lys Arg Leu Ala Val Arg Glu Leu Ala Ala Ala Thr Leu Thr His
                                   676                 <-----S2AR---
GTG CAA GGC TGC AGC GCT GTG GAC GGT CTG GGC CTG CAG CAG GGC CTT GAG CGC
Val Gln Gly Cys Ser Ala Val Asp Gly Leu Gly Leu Gln Gln Gly Leu Glu Arg
-------------                       730                                 757
CTC TAT GAG ATG ATC CTC AAG AGG AAG AAG GCA GCT CGG GGT GGC AAG AAG AGA
Leu Tyr Glu Met Ile Leu Lys Arg Lys Lys Ala Ala Arg Gly Gly Lys Lys Arg
                                   784                                 811
CGG TGA ccc aag ccc ccc ctc cct ttc ctc cca cct agt agg ggt ctg cac act
Arg STOP
                                   838                                 865
tgg aca gca ggg tgg gac cag cct gtg acc tct cag tca gac tgg ggt gca gga
                                   892                                 919
cct gtc cac ctc aat gaa gga gag agg agc atg ggg tgt ccc gtt ttg gtg cca
                                   946                                 973
cac tgg ggt ggg gat ggg aga tgg gat gtc ttt gca tat ctc tct cat cct ctc
                                  1000                                1027
tgg aga agt ggg cgc tgc agg act gtg gag acg taa atg taa act gtg act cta
                                  1054                                1081
cct cga ccc tgt tyc tta ttt ttc ttc tct ggc taa aaa ttt tta att gga tgt
                                  1108                                1135
gtt tgg ggs cgg ggg gat gga agt gac ttg gag aat gtg ttt ggg atg aaa taa
                                  1162                                1189
cta tct ccc ctt cct ctg tcc ccc aac tgg gga gtc tcc cca ggc tgc ttt tct
                                  1216                                1243
agg aat acc agt cac ata gtt ttt att ttt gtg tct gtg aaa gtg cca aga acc
                                  1270                                1297
cct ccc cac att tgt aga tcc atg acc ctt ttt ata agc tgt gtg tgt cct ctg
                                  1324                                1351
tat tat tgt tat taa cta ttt ttt agc att tgc ctg taa gtt att aaa gac tga

taa ctg tag ctc tta aaa aaa aaa a
```

FIG. 1. Nucleotide sequence and predicted protein translation of the human ARF4L gene. ARF4L was initially localized on chromosome 17 by hybridization of a cDNA probe to YACs 19DC6 and 300C2, which form part of a physical contig on 17q12–q21 (1). The localization of ARF4L was confirmed by PCR (using primers S2AF and S2AR), which demonstrated that ARF4L was specifically amplified from genomic clones located on 17q12–q21. Amplification of total human DNA using the same primer pair produced a single band, which demonstrates the specificity of the primers for the ARF4L locus.

```
rARL4    MGNGLSDQTSILSSLPSFQSFHIVILGLDCAGKTTVLYRLQFNEFVNTV  49
                    S LP FQ+ H+V++GLD AGKT++LYRL+F EFV +V
hARF4L   MGNHLTEMAPTASSFLPHFQALHVVVIGLDSAGKTSLLYRLKFKEFVQSV  50
                      + ++++GLD+AGKT++LY+LK  E  ++
hARF1      MGNIFANLFKGLFGKKEMRILMVGLDAAGKTTILYKLKLGEIVTTI  46
                                    PM1     G1

rARL4    PTKGFNTEKIKVTLGNSKTVTFHFWDVGGQEKLRPLWKSYTRCTDGIVFV  99
         PTKGFNTEKI+V LG S+ +TF  WDVGGQEKLRPLW+SY R TDG+VFV
hARF4L   PTKGFNTEKIRVPLGGSRGITFQVWDVGGQEKLRPLWRSYNRRTDGLVFV 100
         PT GFN E +        + I+F VWDVGGQ+K+RPLWR Y + T GL+FV
hARF1    PTIGFNVETVEY-----KNISFTVWDVGGQDKIRPLWRHYFQNTQGLIFV  91
         PM2                      PM3

rARL4    VDSVDVERMEEAKTELHKITRISENQGVPVLIVANKQDLRNSLSLSEIEK 149
         VD+ + ER+EEAK ELH+I R S+NQGVPVL++ANKQD   +LS +E+EK
hARF4L   VDAAEAERLEEAKVELHRISRASDNQGVPVLVLANKQDQPGALSAAEVEK 150
         VD+ + ER+ EA+ EL R+    +       LV ANKQD P A++AAE+
hARF1    VDSNDRERVNEAREELMRMLAEDELRDAVLLVFANKQDLPNAMNAAEITD 141
                                        G2

rARL4    LLAMGELSSSTPWHLQPTCAIIGDGLKEGLEKLHDMIIKRRRKMLRQQKKK 199
          LA+ EL+++T  H+Q   A+ G GL++GLE+L++MILKR+K  R  KK+
hARF4L   RLAVRELAAATLTHVQGCSAVDGLGLQQGLERLYEMILKRKKAARGGKKR 200
         +L +  L       ++Q   A  G GL +GL+ L   +  +K
hARF1    KLGLHSL-RHRNWYIQATCATSGDGLYEGLDWLSNQLRNQK           181
                           G3

rARL4    R                                                   200
         R
hARF4L   R                                                   201
```

FIG. 2. Comparison of the predicted amino acid sequence of the human ARF4L protein with rat ARL4 and human ARF1. The deduced amino acid sequence of ARF4L was aligned with ARL4 and ARF1 with the help of the BLAST algorithm. Hyphens represent gaps introduced into the sequences for optimal alignment; + indicates a conservative substitution. The consensus motifs presumed to be involved in phosphate/magnesium binding (PM1–3) and guanine nucleotide binding (G1–3) are underlined in the ARF1 sequence.

in brain, heart, or muscle (5). The rat ARL4 gene, which shows greatest homology to ARF4L, is not expressed in undifferentiated fibroblasts, but is abundant in differentiated cells with an adipocyte-like phenotype (5). This may suggest that the function of ARL4 is related to the differentiation state of fibroblasts, although the ARL4 gene also appears to be expressed in heart, brain, and muscle. At the moment, therefore, it is difficult to speculate on the function of the human ARF4L protein without further insight into the functions of the other ARF and ARF-like proteins.

ACKNOWLEDGMENTS

We are grateful to Margaret Robertson and Elizabeth Lawrence for DNA sequencing, and Ed Meenen for synthesizing oligonucleotides. S.A.S. is supported by an EMBO fellowship. This work was supported by Grant DAMD17-94-J-4129 from the Department of Defense to R.W.

REFERENCES

1. Albertsen, H. M., Smith, S. A., Mazoyer, S., Fujimoto, E., Stevens, J., Williams, B., Rodriguez, P., Cropp, C. S., Slijepcevic, P., Carlson, M. N., Robertson, M., Bradley, P., Lawrence, E., Harrington, T., MeiSheng, Z., Hoopes, R., Sternberg, N., Brothman, A., Callahan, R., Ponder, B. A. J., and White, R. (1994). A physical map and candidate genes in the BRCA1 region on chromosome 17q12–21. *Nature Genet.* **7:** 472–479.
2. Balch, W. E., Kahn, R. A., and Schwaninger, R. (1992). ADP-ribosylation factor is required for vesicular trafficking between

the endoplasmic reticulum and the cis-Golgi compartment. *J. Biol. Chem.* **267:** 13053–13061.

3. Clark, J., Moore, L., Krasinkas, A., Way, J., Battey, J., Tamkun, J., and Khan, R. A. (1993). Selective amplification of additional members of the ADP-ribosylation factor (ARF) family: Cloning of additional human and Drosophila ARF-like genes. *Proc. Natl. Acad. Sci. USA* **90:** 8952–8956.
4. Kahn, R. A., Randazzo, P., Serafini, T., Weiss, O., Rulka, C., Clark, J., Amherdt, M., Roller, P., Orci, L., and Rothman, J. E. (1992). The amino terminus of ADP-ribosylation factor (ARF) is a critical determinant of ARF activities and is a potent and specific inhibitor of protein transport. *J. Biol. Chem.* **267:** 13039–13046.
5. Schurmann, A., Breiner, M., Becker, W., Huppertz, C., Kainulainen, H., Kentrup, H., and Joost, H.-G. (1994). Cloning of two novel ADP-ribosylation factor-like proteins and characterization of their differential expression in 3T3-L1 cells. *J. Biol. Chem.* **269:** 15683–15688.
6. Serafini, T., Orci, L., Amherdt, M., Brunner, M., Kahn, R. A., and Rothman, J. E. (1991). ADP-ribosylation factor is a subunit of the coat of Golgi-derived COP-coated vesicles: A novel role for a GTP-binding protein. *Cell* **67:** 239–253.
7. Valencia, A., Chardin, P., Wittinghofer, A., and Sauder, C. (1991). The RAS protein family: Evolutionary tree and role of conserved amino acids. *Biochemistry* **30:** 4637–4648.

Am. J. Hum. Genet. 54:516-525, 1994

Genetic Mapping of the BRCA1 Region on Chromosome 17q21

Hans Albertsen,* Rosemarie Plaetke,* Linda Ballard,* Esther Fujimoto,* Judith Connolly,* Elizabeth Lawrence,* Pilar Rodriguez,* Margaret Robertson,*,† Paige Bradley,* Bruce Milner,* David Fuhrman,* Andy Marks,* Robert Sargent,* Peter Cartwright,* Nori Matsunami,*,† and Ray White*,†

*Eccles Institute of Human Genetics and †Howard Hughes Medical Institute, University of Utah, Salt Lake City

Summary

Chromosome 17q21 harbors a gene (BRCA1) associated with a hereditary form of breast cancer. As a step toward identification of this gene itself we developed a number of simple-sequence-repeat (SSR) markers for chromosome 17 and constructed a high-resolution genetic map of a 40-cM region around 17q21. As part of this effort we captured genotypes from five of the markers by using an ABI sequencing instrument and stored them in a locally developed database, as a step toward automated genotyping. In addition, YACs that physically link some of the SSR markers were identified. The results provided by this study should facilitate physical mapping of the BRCA1 region and isolation of the BRCA1 gene.

Introduction

Breast cancer is an often fatal neoplastic disease of mammary tissue that causes ~50,000 deaths/year in the United States alone. Most cases are sporadic, with no apparent genetic lineage. However, a hereditary form of the disease, observed in approximately 5% of all cases and characterized by early age at onset, has been genetically linked to marker loci on chromosome 17q21 (Hall et al. 1990). A summary of 13 published reports located the cancer-predisposing locus, BRCA1, within the interval defined by D17S250 (mfd15) and D17S588 (42D6)(Easton et al. 1993). Several authors, however, had suggested a more narrow localization defined proximally by THRA1 and distally by D17S579 (mfd188) (Bowcock et al. 1993; Devilee et al. 1993; Simard et al. 1993; Smith et al. 1993). The genetic distance separating these two markers is ~5 cM (O'Connell et al. 1993). In the study reported here, we found supporting evidence for this estimate and identified 25 simple-sequence-repeat (SSR) markers, many of which fall within this region or flank it closely on either side. To facilitate efforts to identify the BRCA1 gene itself, we have integrated these new markers with previously known markers to form a highly resolved genetic map of this region of the long arm of chromosome 17.

The new SSR markers were developed as part of a comprehensive effort to generate large numbers of highly informative DNA markers with which a high-resolution genetic map of the entire human genome might be constructed. Such loci are easily typed by the PCR, using unique primers flanking each variable-repeat region (Saiki et al. 1988; Weber and May 1989; Orita et al. 1990). That markers of this type are abundant and highly informative has been shown by several studies describing di-, tri-, and tetranucleotide repeats (Litt and Luty 1989; Economou et al. 1990; Weissenbach et al. 1992; Melis et al. 1993). As part of an effort to automate SSR genotyping, five of our new markers in the BRCA1 region were fluorescently labeled, gel separated, and analyzed on an automated ABI373A DNA sequencer (Ziegle et al. 1992). Specialized software was developed to facilitate capture and storage of allelic information.

In addition to providing genetic information, the new SSR markers can serve as anchor points for a physical map of the BRCA1 region. By screening the CEPH library of YACs (Albertsen et al. 1990), we have identified several YACs that provide evidence for physical

Received July 27, 1993; accepted for publication October 25, 1993.

Address for correspondence and reprints: Dr. Hans Albertsen, Eccles Institute of Human Genetics, Building 533, Room 2100, Salt Lake City, UT 84112.

0002-9297/94/5403-0015$02.00

linkage among some of the SSR markers on the genetic map reported here.

Material and Methods

Preparation and Screening of Sau3A-digested Genomic M13 Library

Genomic DNA from lymphocytes of a human male was digested with *Sau*3AI, partially filled in with Klenow large fragment and dGTP/dATP, and size fractionated on a 1.2% agarose gel. DNA fragments of 400–900 bp were recovered from the gel and ligated with M13mp18 vector that had been digested with *Sal*I and partially filled in with dTTP/dCTP. The library was plated on Luria-Bertani-medium plates, transferred to nylon membranes and UV-cross-linked, and hybridized sequentially with end-labeled oligonucleotide probes $(dG\text{-}dT)_{10}$, $(dA_3\text{-}dT)_6$, $(dA_3\text{-}dG)_6$, and $(dA\text{-}dG\text{-}dA\text{-}dT)_6$. Hybridization was carried out for 3–4 h at 65°C in 10% polyethylene glycol, 7% SDS, and 1.5 × SSPE ($NaCl$–NaH_2PO_4–EDTA buffer). Membranes were washed in 6 × SSC and 0.1% SDS at 65°C.

Sequencing of Positive Clones

Positively hybridizing plaques were directly picked into 100-μl PCR cocktails containing 10 mM Tris-HCl pH 8.8, 40 mM NaCl, 1.5 mM $MgCl_2$, 5 pmol of each vector primer (A—TGT AAA ACG ACG GCC AGT CGC CAG GGT TTT CCC AGT CAC GAC; and B—CAG GAA ACA GCT ATG ACC AGC GGA TAA CAA TTT CAC ACA GGA); 2.5 μmol each of dNTPs, and 2 U of *Taq* DNA polymerase (Boehringer-Mannheim, Indianapolis). The reaction mixtures were heated at 94°C for 6 min, and the PCR was performed on a GeneAmp PCR System 9600 (Perkin-Elmer) for 25 cycles as follows: 95°C for 20 s (denaturation), 60°C for 20 s (annealing), and 72°C for 20 s (extension). The amplified inserts were purified with a Centricon-100 microconcentrater (Amicon, Danvers, MA), and the sequencing reactions were carried out on ABI Catalyst (Applied Biosystems, Foster City, CA) by the dideoxy chain-termination method using fluorescently labeled M13 sequencing primers: −21M13—TGT AAA ACG ACG GCC AGT; and M13RP1—CAG GAA ACA GCT ATG AC. The sequences were collected on an ABI383A sequencer (Applied Biosystems, Foster City, CA).

Primer Synthesis

Primers were synthesized in 40-nmol reactions by using an ABI394 DNA/RNA synthesizer (Applied Biosystems, Foster City, CA). After lyophilization, the primers were resuspended in 300 μl of TE_{-4} (10 mM Tris-Cl pH 7.8, 0.1 mM EDTA pH 8.0). Primer concentrations were determined by densitometry at 260 nm, and working stocks were prepared at 25 pmol/μl.

Primer Selection and Development

PCR primers were developed using a locally developed computer program, OLIGO (J.-M. Lalouel and T. Elsner, personal communication), on the basis of genomic sequence flanking the repeats. This program allows sequence comparisons between oligonucleotide primers and Alu consensus sequences, to detect homologies and to exclude primers with homologies above a user-defined threshold. PCR conditions were optimized with respect to $MgCl_2$ concentration (1.0, 1.25, 1.5, 2.0, 3.0, and 4.0 mM) and annealing temperatures. The size of each PCR product was determined from the sequence and verified, first through gel electrophoresis in 5% NuSieve 3:1 agarose (FMC BioProducts, Rockland, ME) and later through acrylamide gel electrophoresis. The primer sequences and their characteristics are listed in table 1 for each of the 25 new SSR markers on chromosome 17.

Radioactive Genotyping

The primer of each pair that showed the least homology to Alu was labeled with ^{32}P in a kinase reaction. PCR was carried out for a total of 30 cycles. Denaturation was done at 95°C for 6 min in the first cycle and for 10 s in each of the subsequent 29 cycles; annealing was done for 10 s at the temperature specific to each set of primers; and extension was done at 72°C for 20 s. The reactions were carried out in a 96-well Techne MW-2 thermal cycler (Techne, Cambridge). PCR products were mixed with a formamide dye solution (98% formamide, 0.1% xylene cyanol, 0.1% bromophenol blue, and 10 mM EDTA) and heated to 94°C for 4 min. A 3-μl aliquot of each sample was electrophoresed through a 7% denaturing polyacrylamide sequencing gel with 32% of deionized formamide, 5.6 M urea, and 1 × TBE (Tris–boric acid–EDTA buffer). After electrophoresis, the gels were exposed to X-ray film without drying, at −70°C for 12–24 h, with intensifying screens.

Fluorescent Genotyping

As with radioactive genotyping, the primer showing the least homology to Alu was labeled, albeit with a fluorochrome, for detection on an ABI373A sequencing instrument. Of the three different dyes, the blue dye

(FAM) can be incorporated directly onto the 5′ end of an oligonucleotide on the DNA synthesizer; labeling with the yellow and green dyes (TAMRA and JOE) requires synthesis of oligonucleotides with aminolink 2 (400808, ABI) attached 5′, and subsequent addition of the appropriate dye-NHS ester. Unreacted dye is removed by means of a PD-10 column (17-0851-01, Pharmacia) (for technical details, see Genescan 672 Software User's Manual, Appendix D, ABI). Unlabeled oligonucleotides are removed on a purification cartridge, OPC (400771, ABI; User Bulletin 51, ABI).

To establish working conditions for each primer pair, PCR was performed on a few samples of DNA from the CEPH reference panel in a GeneAmp PCR System 9600. Reaction volumes of 100 μl contained 400 ng of template, 200 μM each dNTP, 0.5 μM each primer, 2.7 U of *Taq* DNA polymerase, 0.24 mM spermidine, 10 mM Tris-HCl pH 8.7, and appropriate concentrations (1.0–4.0 mM) of $MgCl_2$. Annealing temperatures were estimated on the basis of primer sequence and were adjusted where necessary. The fluorescent PCR products were also tested for optimal signal on the sequencing instrument, by analyzing aliquots taken at 14, 17, 20, and 23 cycles to determine the number of cycles necessary to observe a specific product with minimal formation of spurious product. Following the establishment of working conditions, the PCR reactions were proportionally scaled down to 25 μl and were used for genotyping individuals in the CEPH panel.

Linkage Analysis

All linkage analyses described in this paper were performed using four programs from the LINKAGE package (version 5.1): CFACTOR, CLODSCORE, CILINK, and CMAP (Lathrop et al. 1984).

Linkage Data

Linkage data used in this study were derived partly from O'Connell et al. (1993), partly from the CEPH database (version 6), and partly from genotypic analysis of new SSR markers that were labeled with ^{32}P or with a fluorochrome.

Identification of YAC Clones

To initiate the development of physical representation from the BRCA1 region, the CEPH library of YACs was screened by PCR according to a protocol described by Green and Olson (1990). Some of the YACs isolated in this way identified close physical linkage for several of the SSR markers described here.

Results and Discussion

To supplement the existing archive of SSR markers, we have developed >2,000 genomic-sequence-tagged markers based predominantly on tetranucleotide repeats, as this type of repeat is, in general, highly informative and tends to show less susceptibility to PCR artifacts such as laddering than the dinucleotide $(CA)_n$ repeats (Litt and Luty 1989; Tautz 1989). To augment the marker density on chromosome 17 specifically, a flow-sorted cosmid library (a gift from Dr. L. Deaven, Los Alamos National Laboratory) was subcloned into the M13 sequencing vector and screened for the presence of selected di- and tetranucleotide repeats; appropriate SSR loci were sequenced for development of primers. Approximately 80 new SSR markers for chromosome 17 were obtained in this way (data not shown). We have now mapped 25 of those markers, by genetic linkage analysis, to a 40-cM region surrounding the BRCA1 locus.

Localization of New SSR Markers to the BRCA1 Region

Strategies to reduce the effort involved in linkage analyses, by reducing the number of genotypes required for map construction, are being implemented in our laboratory. One of these strategies begins by genotyping each new marker in only four CEPH pedigrees (884, 1331, 1332, and 1362) and comparing two-point lod scores obtained with selected loci, to determine an approximate chromosomal location; the marker is then genotyped on a panel of CEPH individuals with known meiotic breakpoints in that region. An initial rough localization of each new SSR marker to the BRCA1 region was based on information from the four selected CEPH pedigrees. The power of a two-point lod score analysis in these pedigrees, when testing SSR markers against the Genethon markers (Weissenbach et al. 1992), depends on the degree of informativeness of both markers in the comparison, as well as on their genetic distance. For two completely informative markers <5 cM apart, the lod scores can approach 20. However, in light of the fact that markers in the comparisons rarely reveal complete genetic information, most SSR markers are assigned to particular chromosomal regions with lod scores of 6–15 and recombination fractions of .05–.20.

Selection of Markers for the BRCA1 Map

The initial stage of the genetic analysis of markers from the BRCA1 region was in part based on results published elsewhere (O'Connell et al. 1993), which had

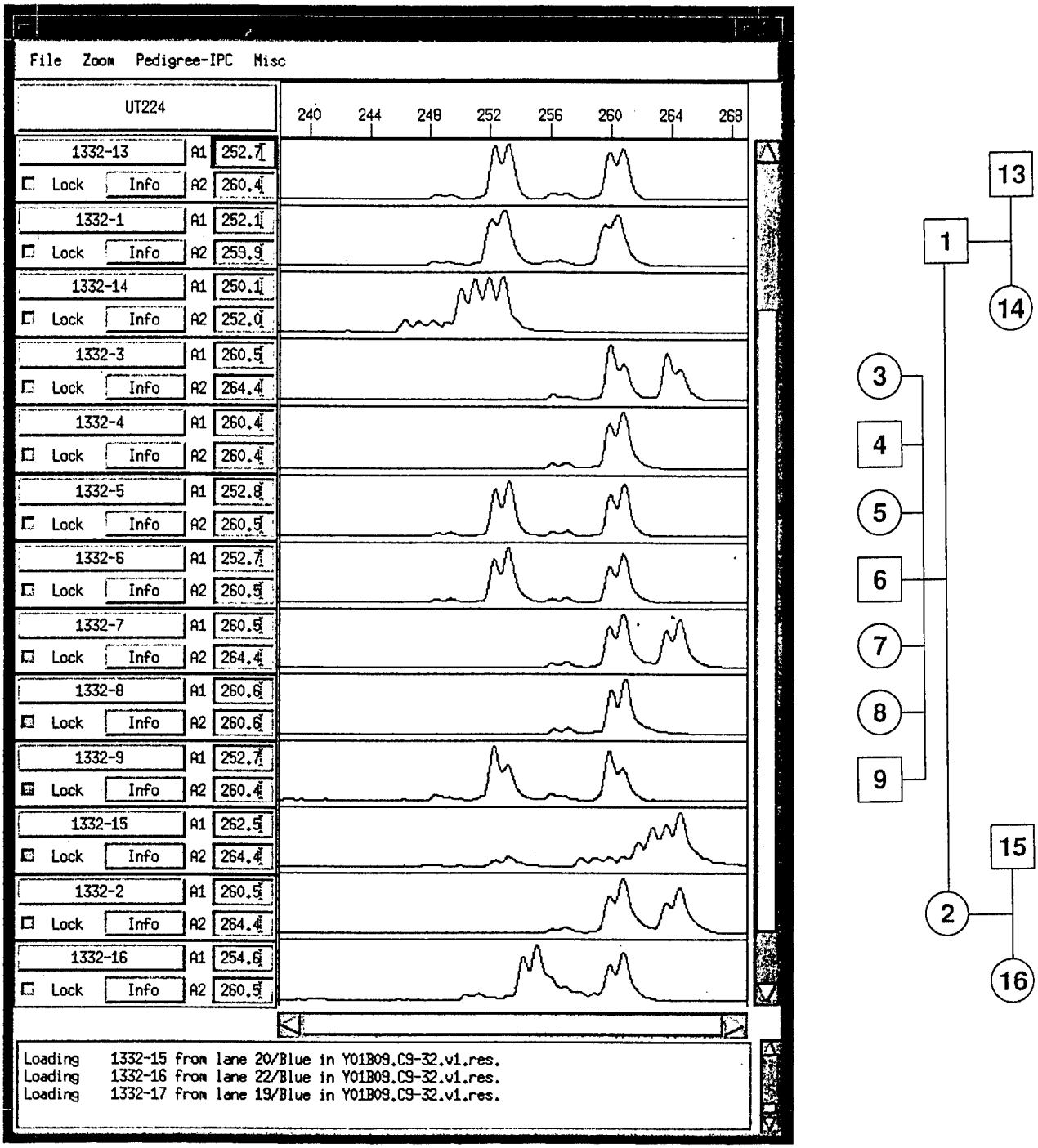


Figure 2 Data traces from fluorescent allelotyping of UT224 labeled with the FAM fluorochrome, as an alternative to traditional visualization of SSR markers with ^{32}P. Each blue trace represents the PCR-amplification product of a single individual who is identified by pedigree number and individual number (e.g., 1332-13) to the left of each trace. Also indicated at left are allele sizes (in base pairs) observed for each individual. The pedigree structure shown to the right is added here to aid interpretation of the traces. The split peaks 1 bp apart, seen in each allele, could be interpreted as an effect of the *Taq* DNA polymerase, which is known to unspecifically add a single A nucleotide to the 3' terminus of an extension product (Clark 1988).

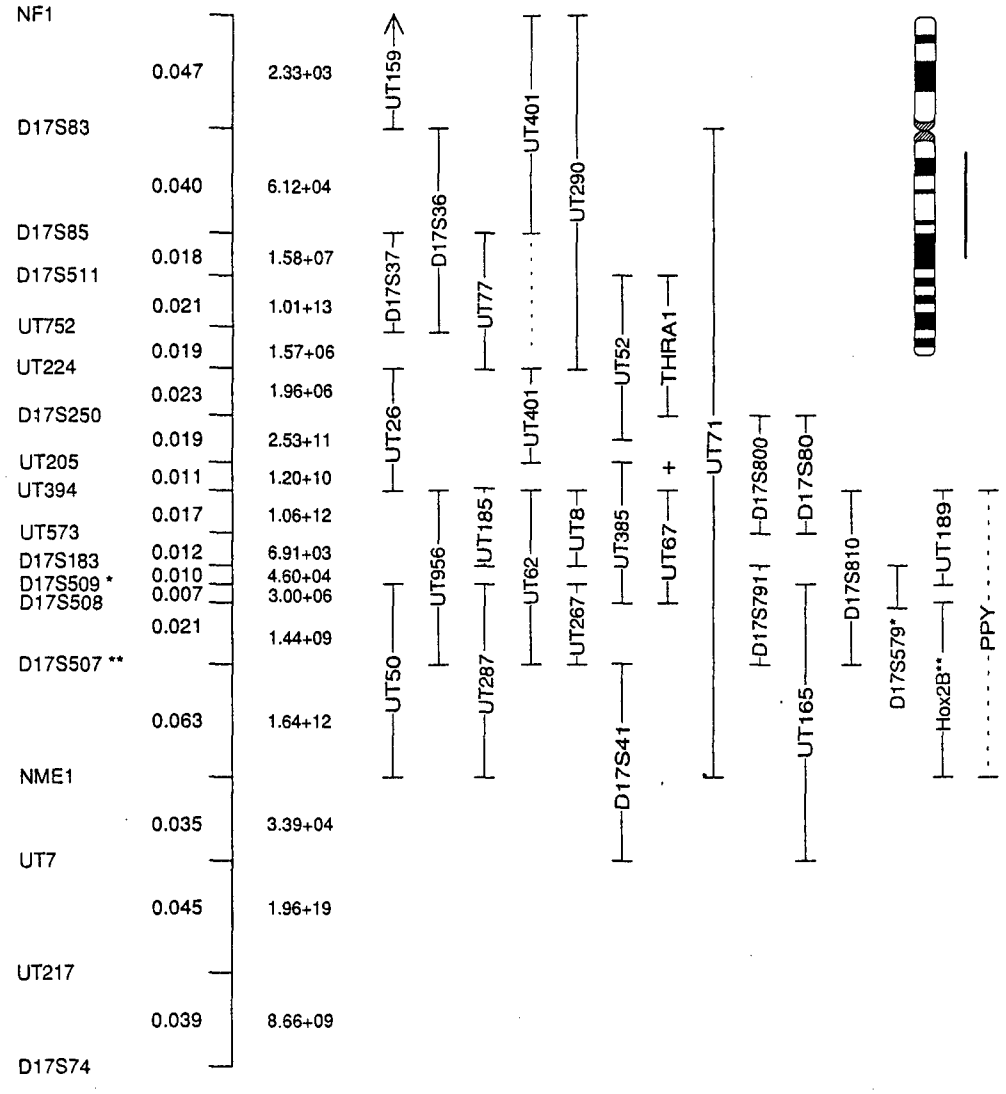


Figure 3 *Upper right,* Karyogram of chromosome 17, with a vertical line to indicate the approximate coverage of the map developed in this study. *Left,* All markers that could be linearly ordered with odds >1,000:1, starting, at the top, with the most centromeric marker, NF1. D17S74 marks the map boundary on the telomeric side. Recombination frequencies between neighboring markers are indicated on the left side of the map, and the odds against inversion in each interval are on the right. The markers that could not be placed in a single interval on the map are placed within confidence intervals of 1,000:1 odds and are illustrated by vertical lines. The two pairs of loci that were haplotyped (see text) are indicated by single and double asterisks (* and **). Marker UT159 is drawn with an arrow pointing up to indicate that the confidence interval extends beyond NF1. Marker UT401 is drawn in two intervals connected by a dashed line to indicate a noncontinuous confidence interval; however, the location of UT401 around D17S250 is favored with >40:1 odds over the other location. The location of THRA1, shown by its confidence interval, is based on the genetic information; however, since this location is in disagreement with observations made with YACs and pulsed-field gel electrophoresis, its physical location is indicated by a plus sign (+). PPY could not be placed into an unambiguous confidence interval with 1,000:1 odds and is therefore shown on the map by a dashed line to indicate its most likely location.

analyzed for errors and, where possible, were tested by secondary typings (see O'Connell et al. 1993). Inconsistency between the genetic and the physical map locations of THRA1 suggested that the genetic data were probably incorrect, in that this gene had been physically linked to UT205 on YAC 44F8 (280 kb) (and on other YACs; data not shown); yet the genetic analysis excluded UT205 from the 1,000:1-odds interval of THRA1. The physical location of THRA1 in the immediate vicinity of the UT205 locus is supported by observations using pulsed-field gel electrophoresis (Lemons et al. 1990). We attempted to place another marker,

breast-ovarian cancer gene (BRCA1) on chromosome 17q21. Am J Hum Genet 52:718–722

Clark JM (1988) Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. Nucleic Acids Res 16:9677–9686

Cropp C, Liderau R, Campbell G, Champene HM-H, Callahan R (1990) Loss of hererozygosity on chromosome 17 and 18 in breast carcinoma: two new regions identified. Proc Natl Acad Sci USA 87:7737–7741

Devilee P, Cornelis RS, Bootsma A, Bardoel A, van Vliet M, van Leeuwen I, Cleton FJ, et al (1993) Linkage to markers for the chromosome region 17q12-q21 in 13 Dutch breast cancer kindreds. Am J Hum Genet 52:730–735

Easton DF, Bishop DT, Ford D, Crockford GP, The Breast Cancer Linkage Consortium (1993) Genetic linkage analysis in familial breast and ovarian cancer: results from 214 families. Am J Hum Genet 52:678–701

Economou EP, Bergen AW, Warren AC, Antonarakis SE (1990) The polyadenylate tract of Alu repetitive elements is polymorphic in the human genome. Proc Natl Acad Sci USA 87:2951–2954

Green ED, Olson MV (1990) Systematic screening of yeast artificial-chromosome libraries by use of the polymerase chain reaction. Proc Natl Acad Sci USA 87:1213–1217

Green P, Helms C, Cartinhour S, Weiffenbach B, Stephens K, Keith T, Bowden DW, et al (1987) Towards a primary genetic linkage map of the human genome: a progress report on linkage studies with more than 300 new RFLPs. Cytogenet Cell Genet 46:623

Hall JM, Lee MK, Newman B, Morrow JE, Anderson LA, Huey B, King M-C (1990) Linkage of early-onset familial breast cancer to chromosome 17q21. Science 250:1684–1689

Joslyn G, Carlson M, Thliveris A, Albertsen H, Gelbert L, Samowitz W, Groden J, et al (1991) Identification of deletion mutations and three new genes at the familial polyposis locus. Cell 66:601–613

Lathrop GM, Lalouel J-M, Ott J (1984) Strategies for multilocus linkage analysis in humans. Proc Natl Acad Sci USA 81:3443–3446

Lemons RS, Eilender D, Waldman RA, Rebentisch M, Frej A-K, Ledbetter DH, Willman C, et al (1990) Cloning and characterization of the t(15;17) translocation breakpoint region in acute promyelocytic leukemia. Genes Chromosom Cancer 2:79–87

Litt M, Luty JA (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am J Hum Genet 44:397–401

Melis R, Bradley P, Elsner T, Robertson M, Lawrence E, Gerken S, Albertsen H, et al (1993) Polymorphic SSR (simple-sequence-repeat) markers for chromosome 20. Genomics 16:56–62

Nakamura Y, Lathrop M, O'Connell P, Leppert M, Barker D, Wright E, Skolnick M, et al (1988) A mapped set of DNA markers for human chromosome 17. Genomics 2:302–309

O'Connell P, Albertsen H, Matsunami N, Taylor T, Hundley JE, Johnson-Pais TL, Reus B, et al (1994) A radiation hybrid map of the BRCA1 region. Am J Hum Genet 54:526–534 (in this issue)

O'Connell P, Plaetke R, Matsunami N, Odelberg S, Jorde L, Chance P, Leppert M, et al (1993) An extended genetic linkage map and an "index" map for human chromosome 17. Genomics 15:38–47

Orita M, Suzuki Y, Sekiya T, Hayashi K (1990) Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction. Genomics 5:874–879

Saiki RF, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Erlich HA (1988) Primer directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487–491

Sheer D, Sheppard DM, le Beau M, Rowley JD, San Roman C, Solomon E (1985) Localization of the oncogene c-erbA1 immediately proximal to the acute promyelocytic leukemia breakpoint on chromosome 17. Ann Hum Genet 49:167–171

Simard J, Feunteun J, Lenoir G, Tonin P, Norman T, The VL, Vivier A, et al (1993) Genetic mapping of the breast-ovarian cancer syndrome to a small interval on chromosome 17q12-21: exclusion of candidate genes EDH17B2 and RARA. Hum Mol Genet 2:1193–1199

Smith SA, Easton DF, Ford D, Peto J, Anderson K, Averill D, Stratton M, et al (1993) Genetic heterogeneity and localization of a familial breast-ovarian cancer gene on chromosome 17q12-q21. Am J Hum Genet 52:767–776

Takeuchi T, Gumucio D, Eddy R, Meisler M, Minth C, Dixon J, Yamada T, et al (1985) Assignment of the related pancreatic polypeptide (PPY) and neuropeptide Y (NPY) genes to regions on human chromosomes 17 and 7. Cytogenet Cell Genet 40:759

Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res 17:6463–6471

Varesco L, Caligo MA, Simi P, Black DM, Nardini V, Casario L, Rocchi M, et al (1992) The NM23 gene maps to human chromosome band 17q22 and shows a restriction fragment length polymorphism with BglII. Genes Chromosom Cancer 4:84–88

Weber JL, May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet 44:388–396

Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, Millasseau P, Vaysseix G, et al (1992) A second-generation linkage map of the human genome. Nature 359:794–801

Ziegle JS, Su Y, Corcoran KP, Nie L, Mayrand E, Hoff LB, McBride LJ, et al (1992) Application of automated DNA sizing technology for genotyping microsatellite loci. Genomics 14:1026–1031

A physical map and candidate genes in the *BRCA1* region on chromosome 17q12–21

H. M. Albertsen[1], S. A. Smith[1], S. Mazoyer[2], E. Fujimoto[1], J. Stevens[1], B. Williams[1], P. Rodriguez[1], C. S. Cropp[3], P. Slijepcevic[2], M. Carlson[1], M. Robertson[1], P. Bradley[1], E. Lawrence[1], T. Harrington[2], Z. Mei Sheng[3], R. Hoopes[4], N. Sternberg[4], A. Brothman[1], R. Callahan[3], B. A. J. Ponder[2] & Ray White[1]

We have constructed a physical map of a 4 cM region on chromosome 17q12–21 that contains the hereditary breast and ovarian cancer gene *BRCA1*. The map comprises a contig of 137 overlapping yeast artificial chromosomes and P1 clones, onto which we have placed 112 PCR markers. We have localized more than 20 genes on this map, ten of which had not been mapped to the region previously, and have isolated 30 cDNA clones representing partial sequences of as yet unidentified genes. Two genes that lie within a narrow region defined by meiotic breakpoints in *BRCA1* patients have been sequenced in breast cancer patients without revealing any deleterious mutations. These new reagents should facilitate the identification of *BRCA1*.

[1]*Eccles Institute of Human Genetics and Howard Hughes Medical Institute, University of Utah, Salt Lake City, Utah 84112, USA*
[2]*CRC Human Cancer Genetics Research Group, Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB21QP, UK*
[3]*Laboratory of Tumor Immunology and Biology, National Cancer Institute, NIH, Bethesda, Maryland 20892, USA*
[4]*The Du Pont Merck Pharmaceutical Company, Wilmington, Delaware 19880-0328, USA*

Correspondence should be addressed to H.M.A.

At the end of 1990, a susceptibility gene for hereditary breast and ovarian cancer, *BRCA1*, was assigned by genetic linkage to the long arm of chromosome 17 (ref. 1). About 5% of all breast cancers appear to be hereditary. Using meiotic breakpoints, a consortium of investigators subsequently refined the region containing *BRCA1* to a ~12 centiMorgan (cM) (sex averaged) interval flanked proximally by *D17S250* and distally by *D17S588* (ref. 2). Other recombination events in *BRCA1* families narrowed the location of *BRCA1* further to a region of less than 4 cM defined by the thyroid receptor gene A1 (*THRA1*) (ref. 3) and *D17S579* (ref. 4). In addition to its role in familial breast and ovarian cancer, *BRCA1* is probably involved in non-familial forms of the disease, a concept which is supported by the observation that 17q is a frequent site of loss of heterozygosity (LOH) in sporadic breast and ovarian tumours.

High-density maps of polymorphic markers and genes in chromosome 17q12–21 have been constructed using genetic[5,6], radiation hybrid[7,8] and FISH[9] data for the purpose of localizing *BRCA1* more precisely. Using the available markers and genes, we have developed a physical map of the region. In turn, we have used the physical map reagents to screen cDNA libraries. In this way we have identified several new genes in the minimal region of 1–2 cM, that is now thought to contain *BRCA1*. These genes are currently being investigated as candidates for the *BRCA1* gene itself.

Construction of the physical map

To develop physical coverage of the *BRCA1* region, large-insert yeast artificial chromosome (YAC)[10,11,12,13] and P1[14] libraries were initially screened with a small collection of polymorphic and non-polymorphic PCR-based markers known to flank *BRCA1*. Several clones were identified that formed small islands of overlapping DNA fragments. With the map seeded in this way, the islands were extended and adjacent islands bridged by walking. This was done by cloning and sequencing the extremities of YAC and P1 inserts from which we could develop new sequence-tagged-site (STS) markers. We used the STS markers to isolate new clones, thereby extending the physical coverage of the region. In the process of assembling the contig, we were careful to test the chromosomal origin of each newly developed pair of primers, to avoid introducing errors in the map as a consequence of chimaeric YAC clones. As Fig. 1 (column 4) indicates, 46 of 127 primer pairs (37%) developed from YAC ends were chimaeric and therefore not useful as map reagents. In all, a total of 112 PCR markers were mapped, of which 92 were developed in our laboratory and 20 were obtained from other sources. Of the 104 YAC clones we analysed, 20 had been identified in parallel from the CEPH YAC library by other research groups who made the mapping information available via World Wide Web (WWW) database servers, and nine clones were isolated from the St. Louis YAC library and made available by the Michigan Genome Center. A total of 33 P1 clones were isolated from the Du Pont P1 phage library. A chart indicating which genomic clones were positive for which markers is shown in Fig. 1.

Several regions of the contig appear to contain YACs which show evidence of rearrangement or internal deletion. For instance, three clones (251H5, 883E6 and 963B10) indicate a link between *D17S579* and PPY, and this link is supported by pulsed-field gel electrophoresis (PFGE) experiments[15]. However, we have not been able to resolve this link unambiguously, most likely because of

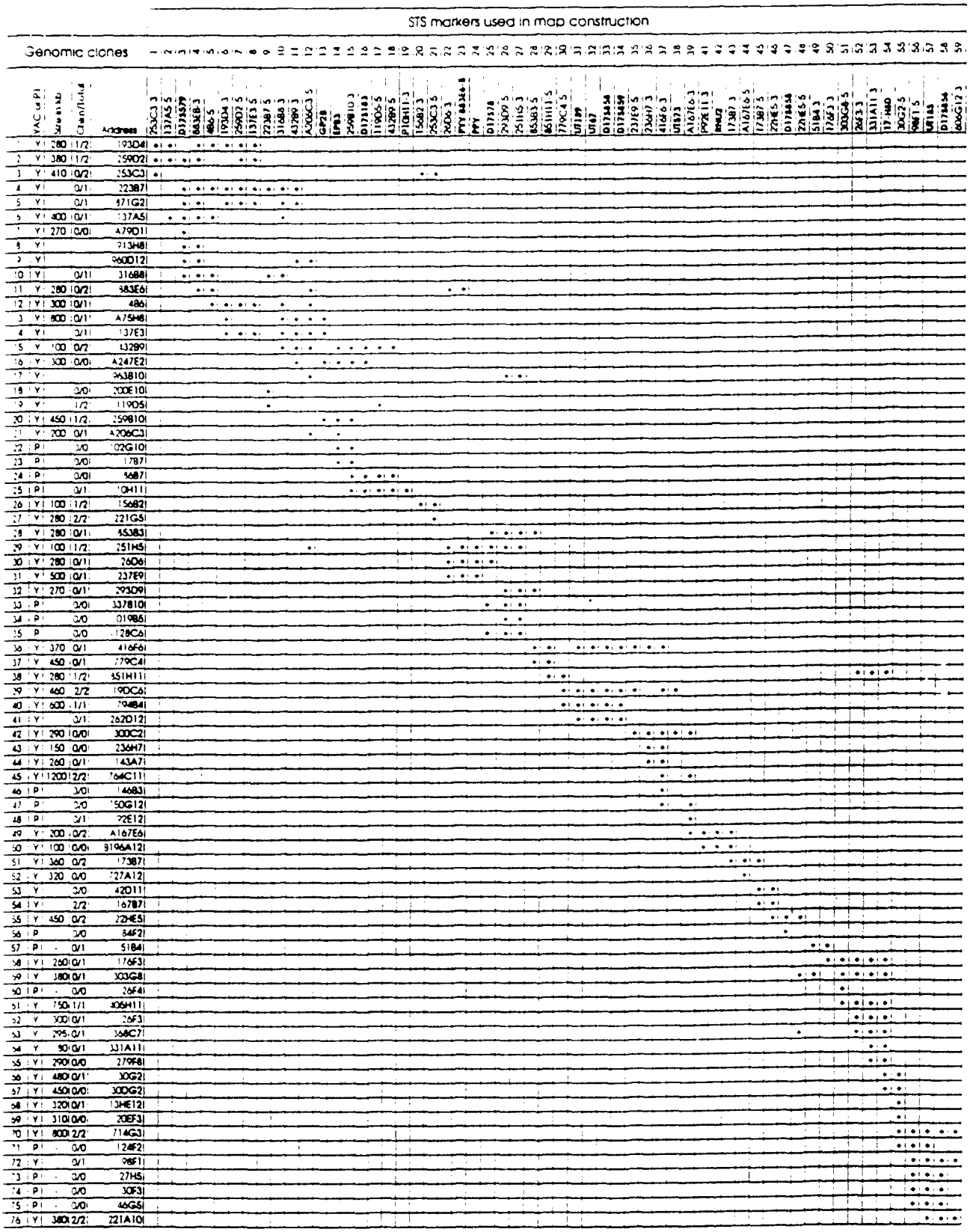
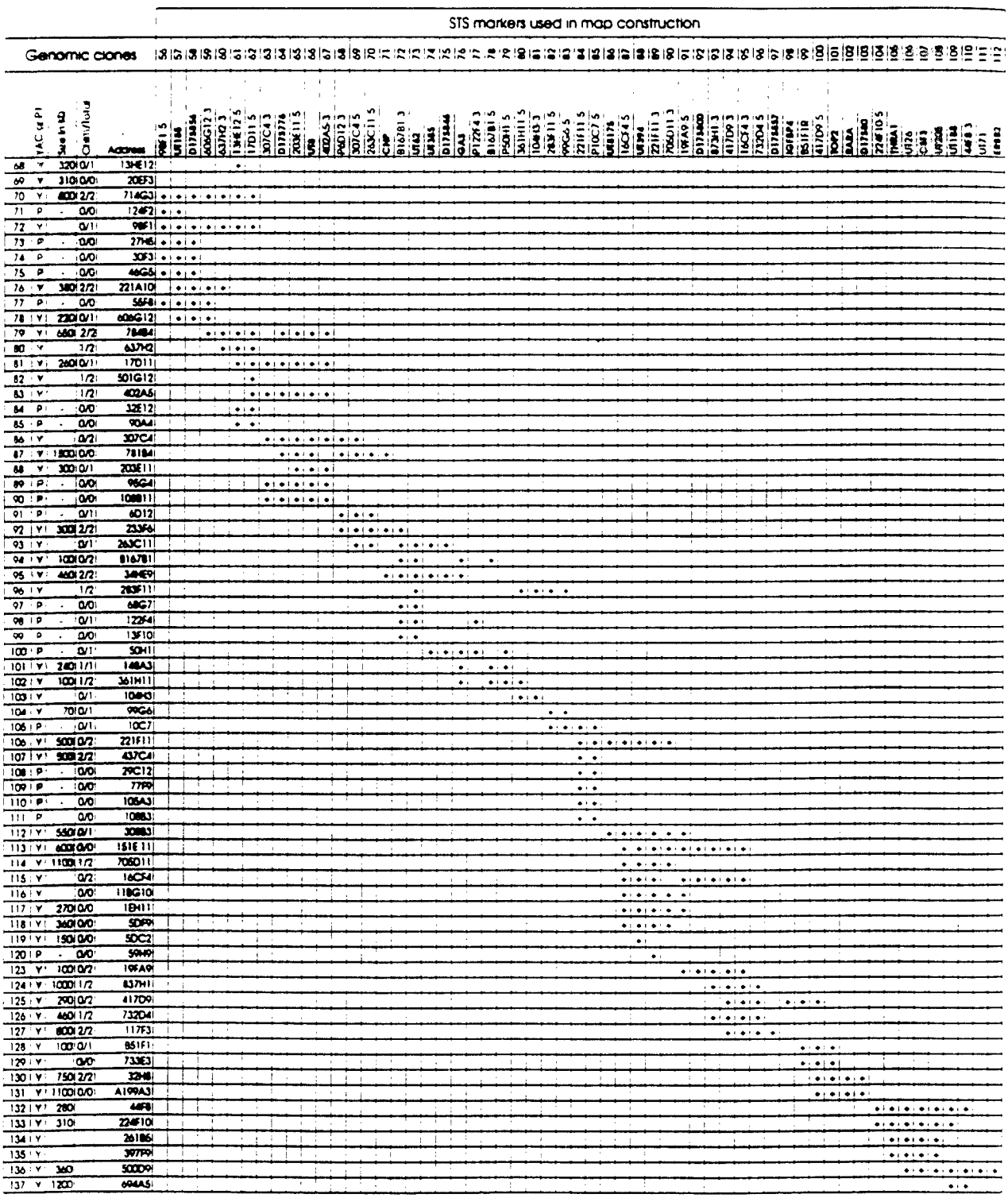


Fig. 1 (This and next page) YAC and P1 clones in the *BRCA1* region. *a* and *b*. Slightly overlapping charts of YAC and P1 clones in the interval between *D17S579* and *ERBB2*. The map indicates which clones, indicated in the address column, are positive for the STS markers, polymorphisms and genes indicated across the top of the Fig. The STS markers, genes and polymorphisms used to construct Fig. 2 are indicated in bold type. Markers that are present in a clone are indicated with a +. Clones from the ICI library can be identified by two letters together within their names. Clones from the St. Louis library have a single letter prefix. The sizes of P1 clones were not determined as they are

internal deletion within YACs in this region. Similarly, genetic evidence supports a close link between *D17S579* and *D17S509* (ref. 6), but it has not been possible to confirm this link with certainty in the YACs. For instance, *D17S579* and *D17S509* identified the same coordinates in the YAC library, 259D2, but it was not possible to recover a single YAC clone positive for both markers even upon testing 40–50 colonies from that plate address. Similar evidence for instability in YAC clones was observed in other regions as well. For example, one address, 99G6, was positive for the markers UT62, 361H11-3, 104H3-3 and 283F11-5. Only upon screening with 283F11-5 was a positive YAC recovered and that YAC was approximately 80 kb in size. The small size of this YAC was consistent with it being considerably shortened by deletion. In several regions of the contig it was only possible to bridge two adjacent STS markers by isolating P1 clones. For instance, PFGE experiments indicated that 283F11-5 and 221F11-5 are in close proximity; however, neither the CEPH nor the ICI YAC libraries contained clones bearing both markers. In contrast, a P1 clone, 10C7, identified with 221F11-5, was positive with both markers thereby bridging the gap. The fragment of DNA represented by 10C7, which was apparently absent from the YAC libraries, appeared to be



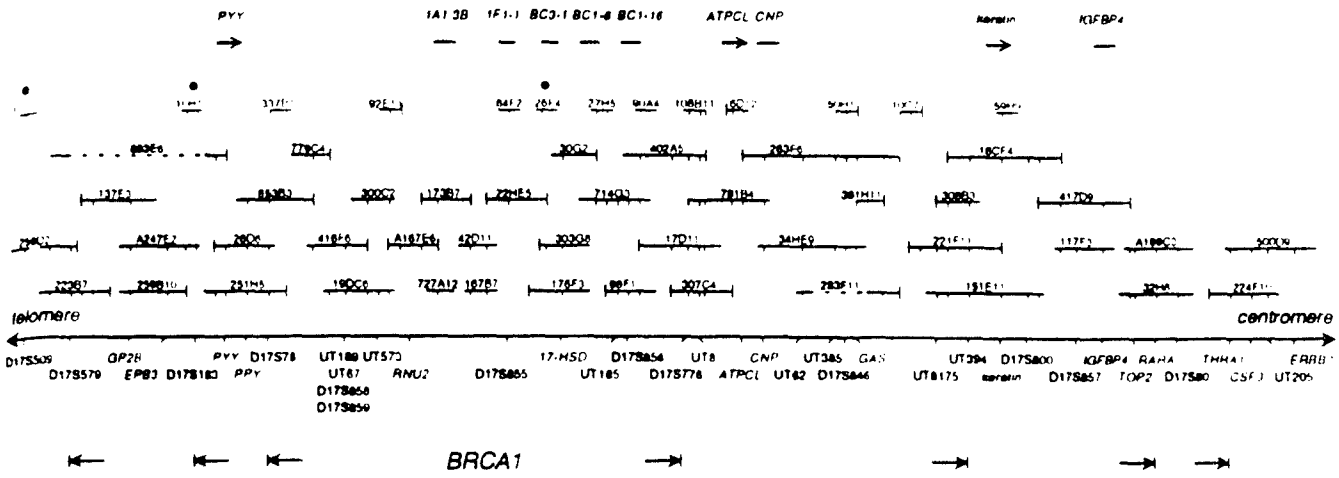


Fig. 2 Contig of overlapping YAC (dark blue) and P1 (light blue) clones from the *BRCA1* region, showing approximate locations of 39 genetic markers (in black type) and genes (in red). Black arrows at the bottom of the map indicate meiotic boundaries that incrementally have helped to refine the *BRCA1* locus. The map is drawn with a minimal set of reagents, composed of 41 YACs and 11 P1 clones, that cover this interval; detailed information about these and additional clones can be found in Fig. 1. The map is estimated to span a physical distance of ≥ 3.5 Mb as judged by the accumulated size of non-overlapping clones. At the top of this Figure, genes and cDNAs that were identified and/or precisely mapped in the present study are shown in red, with the direction of transcription indicated, where known. Circles indicate the three P1 clones mapped by FISH in the experiment shown in Fig. 3, in their respective colours. A previous candidate for the *BRAC1* gene, *MDC*, was identified by Emi *et al.*[40] and localized between *D17S579* and *PPY*.

amenable to cloning in the P1 vector.

A newly identified polymorphism, UT8175, recognizing locus *D17S1136*, was developed from the end-clone of YAC 308B3, which contained a nearly perfect $(A_3G)_{15}$ repeat. Primers designed from this locus revealed four alleles and genetic analysis in a subset of CEPH pedigrees confirmed regional localization by showing tight linkage to UT394 and *D17S800*.

A map of selected overlapping clones from the 4 cM region between D17S579 and *THRA1* is shown in Fig. 2. The information on the order of the markers is derived from the overlapping P1 and YAC clones (Fig. 1). The region between *PPY* and *D17S579*, which appears to contain YACs that have suffered deletions or rearrangements, has been included in our map. A small break in the contig occurred between *THRA1* and *RARA*. No YACs could be isolated that linked *THRA1* to the contig, but a link between *THRA1* and *RARA* has been demonstrated by PFGE[16]. Since a significant proportion of the YACs contain deletions, and only about 100 points along the length of the map have been tested, we cannot exclude the possibility that other small deletions exist which remain undetected. Overall, the map establishes the likely linear order of a large collection of PCR markers over a ~4 megabases (Mb) region.

Table 1 Results of dual-colour P1 hybridization

P1 triplet	Signal between (%)	Signal superimposed (%)	Signal outside (%)
110D12–<u>10H11</u>–26F4	81	7	12
110D12–<u>92E11</u>–124F2	65	22	13
10H11–<u>124F2</u>–50H1	65	15	20
26E4–<u>108B11</u>–59H9	79	16	4
26E4–<u>50H1</u>–59H9	90	6	5
108B11–<u>50H1</u>–59H9	92	4	4

Two of the P1 clones were labelled with a red fluorochrome and a third (underlined), was labelled with a green fluorochrome. For each experiment 100 interphase chromosomes were scored. The Table indicates the percentage of chromosomes in which the green-labelled clone hybridized between the two red labelled clones (column 2), superimposed on one of the red-labelled clones (column 3), or outside of the interval defined by the two red labelled clones (column 4).

Verification of the map by *in situ* hybridization

To help verify our physical map, we analysed several YAC and P1 clones by *in situ* hybridization. Four non-chimaeric YACs (A79D1, 397F9, 16CF4 and 17D11) that were ascertained with *D17S579*, *THRA1*, UT394 and UT8 respectively, were each tested on 20 metaphase spreads and shown to map to the proximal region chromosome 17q. Dual-colour hybridization of the YACs further supported the relative order suggested by PCR experiments (data not shown). In another series of experiments, the relative order of selected triplets of eight P1 clones (110D12, 10H11, 92E11, 26F4, 124F2, 108B11, 50H1 and 59H9) was determined by hybridization to interphase preparations, and found to be in agreement with the physical map (Table 1). An example of one such experiment is shown in Fig. 3, which confirmed that clone 10H11 is located between 110D12 (ascertained with D17S509) and 26F4.

Identification of genes in the *BRCA1* region

Crucial to the identification of *BRCA1* is the isolation of cDNA clones from the minimal region that is thought to contain the disease locus. The minimal region has been systematically narrowed by the identification of critical recombination events in *BRCA1* families; the most recent observations suggest a location for *BRCA1* proximal to *D17S78* (ref. 17) and distal to *D17S776* (ref. 18), a genetic distance of 1–2 cM. Using YAC and P1 clones from this region, we have located and determined the nucleotide sequence of five genes (Table 2 and Fig. 2), and have identified at least 30 other cDNA clones for which we have partial sequence information. In one series of experiments, YACs 727A12 and 19DC6 were isolated from the

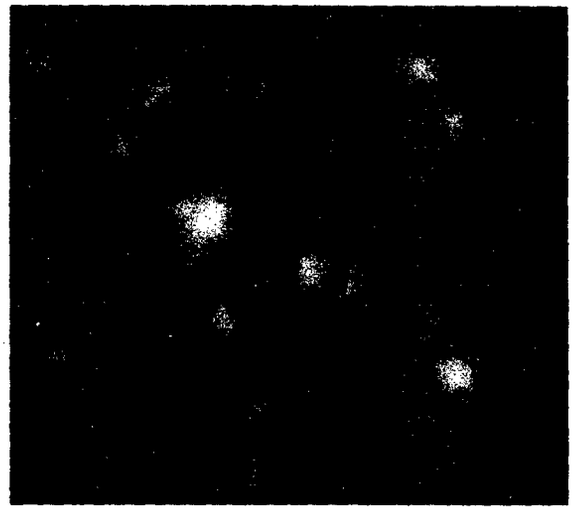


Fig. 3 Fluorescent *in situ* hybridization of P1 clones In this example, three P1 clones (110D12, 10H11 and 26F4), from the *BRCA1* region on the proximal long arm of chromosome 17 were mapped relative to each other. This experiment performed on F6606 skin fibroblast cells arrested at interphase, confirmed that 10H11 (green) is located between 110D12 and 26F4 (both red).

endogenous yeast chromosomes by PFGE, radio-labelled, and used as hybridization probes to screen a fetal-brain cDNA library. YAC 727A12 identified several positive clones, one of which, 1G1-1, had sequence identity with 1A1.3B, which was recently reported by Campbell and colleagues[19]. YAC 19DC6 identified more than 20 positive clones, although most of these clones appeared to be derived from the chimaeric ends of the YAC. As many of the YACs in the critical region had chimaeric ends and YAC probes proved troublesome to radiolabel with sufficient specific activity, we used the P1 clones as probes whenever possible. For example, P1 clones 51D3 and 27H5 identified 15 cDNA clones, representing three different genes. The nucleotide sequence of one of these clones, BC1-16, showed strong homology to the Rab5 family of genes[20], and the translated sequence of another clone, BC3-1, showed sequence identity to the Ki antigen[21]. A third clone, BC1-6, had partial homology to the yeast transcriptional factor, GCN5 (ref. 22), although the start codon in the human gene has not yet been localized with certainty. Work is in progress to map precisely and to determine the complete nucleotide sequence of all clones that have been isolated.

Table 2 Genes identified between *D17S579* and *THRA1*

Clone name	Genbank no.	Homology to: Gene name	Clone ID	% identity
883E6-5 (STS)	L18244	*PYY*	L25648	98%
1G1-1	—	1A1.3B	X76952	100%
1F1-1	not submitted[a]	rat L21	A33295	92%
BC3-1	V11292	*Ki* antigen	A60537	100%
BC1-6	not submitted[a]	*GCN5*	X68628	41%
BC1-16	V11293	dog *Rab5c*	S38625	89%
263C11-5 (STS)	L18219	*ATPCL*	X64330	100%
221F11-3 (STS)	L18207	*HSHB2A_1*	X63337	81%

[a]Submission pending completion of sequence.

Other genes on the physical map

We placed a number of known genes which had been mapped genetically on 17q12–21 accurately on our physical map (Fig. 2) by designing STS primers from the published sequences. For example, the 2'-3'-cyclic nucleotide 3'-phosphohydrolase (*CNP*) gene[23], which had previously been mapped in the 6-cM interval between *THRA1* and nerve growth factor receptor (*NGFR*), has now been localized between *D17S776* and *GAS*. Similarly, the insulin-like growth factor-binding protein 4 (*IGFBP4*) gene[24] which had been mapped on 17q12–21, was localized between *D17S857* and *RARA*.

Other genes were localized through comparisons of the sequences of the YAC end clones with published sequence data. Such comparisons revealed that the 5' end of YAC 417D9 had identity to part of the human DNA topoisomerase II (*TOP2*) mRNA[25] (GenBank accession no. J04088). The sequence identity extended for 79 basepairs (bp) and diverged at position 4304 in the J04088 sequence, suggesting an intron/exon boundary at this position. The orientation of the coding sequence of the *TOP2* gene in the YAC indicates that transcription of this gene is toward the centromere. Similarly, the 5' end of YAC 263C11 had 50 bp of identity to the human ATP citrate lyase (*ATPCL*) gene[26] (GenBank accession no. X64330), suggesting that *ATPCL* is also located within the *BRCA1* region. Divergence of the two sequences at position 1556 in the X64330 sequence suggested that an exon/intron boundary is located at this position. The orientation of the coding sequence in the YAC end-clone indicated that transcription of this gene also is toward the centromere. The 5' end of 883E6 showed 98% identity over its entire length to 398 bp of the 3' untranslated region of the peptide YY (*PYY*) gene[27] (GenBank accession no. L25648), indicating that *PYY* is located close to the pancreatic polypeptide Y (*PPY*) gene. *PYY* and *PPY* are members of the small-peptide hormone family. The direction of transcription of *PYY* is also toward the centromere. Finally, the DNA sequence of the 3' end clone from YAC 211F11 showed 85% sequence similarity to the sheep high-sulphur keratin genes B2A and B2D. Translation of the end clone sequence revealed an 81% similarity on the amino acid level to the human keratin gene *HSHB2A_1* (Genbank accession no. X63337), suggesting that we have identified a previously unknown keratin gene, or a pseudogene, that belongs to the keratin gene cluster located on 17q. This gene, if expressed, is transcribed towards the centromere.

Mutation screening in candidate genes

The coding regions of two of the newly identified genes, BC1-16 and the Ki antigen, were screened for mutations by direct sequencing of cDNA from 24 unrelated patients. Of these patients, 11 came from confirmed *BRCA1* families and 13 were selected on the basis of an unusually high incidence of breast and/or ovarian cancer among their first-degree relatives. The analysis revealed several silent basepair substitutions in both genes, but no missense or frame-shift mutations that could characterize either gene as *BRCA1*.

Discussion

We have described the construction of a physical map on chromosome 17q, between *THRA1* proximally and *D17S579* distally, that encompasses the *BRCA1* locus. The

map comprises 137 overlapping YAC and P1 clones onto which we have ordered 112 PCR markers. We estimate that our map spans ~4 Mb of DNA.

In many places, attempts to construct a physical contig with YACs highlighted the need for an alternative cloning system. Several parts of the map were not represented in YACs, as a result of small deletions or rearrangements which presumably occurred during construction of the YAC library. Other YACs showed evidence of instability, losing specific markers upon replica-plating. In most cases, however, these problems were overcome by supplementing the YAC map with clones from a P1 library. For example, it was not possible to obtain a YAC clone that was positive for both 283F11-5 and 221F11-5 from any of three YAC libraries, although both markers were present on a single 80 kb P1 clone. These experiments demonstrated that investigators should be cautious in interpreting mapping results, and whenever possible employ more than a single source of DNA fragments.

Since the original observation of King and colleagues that a breast and ovarian cancer susceptibility gene, *BRCA1*, was linked to *D17S74* (ref. 1), the localization of *BRCA1* on 17q has been incrementally refined by several groups on the basis of meiotic recombination events in *BRCA1*-linked families[2–4,17,18,28,29]. The most recent observations suggest a localization distal to *D17S776* and proximal to *D17S78*, a distance of 1–2 cM, which is completely contained within the physical contig described here.

Analysis of loss of heterozygosity (LOH) in breast tumours defined a different minimal region than the one defined by meiotic breakpoints. Since *BRCA1* is widely thought to be a tumour suppressor gene[30], the identification of the smallest region of deletion in sporadic breast and ovarian tumours has received much attention in defining new boundaries for the localization of *BRCA1*. From an analysis of 130 breast tumours, Cropp *et al.*[31] have demonstrated a minimal region of LOH centered around *D17S846*, which is approximately 0.5 Mb centromeric of *D17S776*. Given the high likelihood that the recombination event placing *BRCA1* telomeric to *D17S776* is genuine, the simplest explanation for these findings is that at least two separate loci on 17q12–21 are important in breast cancer development: *BRCA1* and a second, more proximal, locus defined by LOH in tumours.

We have mapped a total of seven genes in the minimal *BRCA1* region defined by meiotic breakpoints. At least one of these genes, *EDH17B2* (17-HSD), has been excluded as *BRCA1* by sequencing the gene in a large number of patients who are members of linked families[17,28]. Another gene, 1A1.3B, was excluded in one study[19], but probably warrants further investigation. Two of the remaining genes, BC1-16 and Ki, have been excluded as candidates in this study. *BC1-16* shows 86% homology at the amino-acid level to the Ras-related gene *RAB5*. *RAB5* is a GTP-binding protein that appears to be located in epithelial cells at the cytoplasmic surface of the plasma membrane[20]. Ki encodes the Ki antigen, which is a highly conserved nuclear protein originally detected with sera from patients with the autoimmune disease, systemic lupus erythematosus (SLE)[21]. We have determined the complete coding sequences of both genes in samples from 24 patients and have found no mutations. The other genes (BC1-6, 1F1-1 and *RNU2*) have not yet been screened for mutations in patient samples, and thus remain untested candidates for *BRCA1*. In addition, some 30 other partial cDNA clones that map in the minimal region, await characterization and testing in the patient samples.

The cDNAs and genomic clones we describe here are important resources for continued characterization and isolation of candidate genes, and should facilitate the identification of *BRCA1* in the near future.

Methodology

Libraries. YAC clones were isolated from three libraries: the CEPH library[16]; the CEPH mega-YAC library[11] and the ICI library[12]. Local copies of the CEPH libraries were screened using a PCR-based strategy[32]. YACs from the ICI library were obtained through a screening service supported by the UK Human Genome Mapping Project. A small number of clones obtained from the St. Louis YAC library[13] had been identified at the Michigan Genome Center. A P1 phage library (DMPC-HFF#1 Series B) containing three human haploid genome equivalents was obtained from Du Pont Merck Pharmaceutical Company[14]. This library was received in the format of 125, 96-well microtitre dishes, with each well containing 12 different clones. The P1 library was arrayed onto nine Hybond N+ (Amersham) filters using a Beckman High Density Replicating Tool. The cells were grown and lysed following standard procedures and the DNA was bound to the filters by UV cross-linking. DNA probes for screening the P1 library were obtained by PCR in which a radio-nucleotide, α-^{32}P-dCTP, was incorporated into the reaction product. Positive P1 addresses were identified through a primary hybridization. Single positive clones were identified by a secondary filter hybridization of an inoculum from the address. P1 DNA was isolated as described[14].

DNA preparation. Agarose plugs of yeast cells containing total YAC DNA (final density of 1.3×10^8 cells ml^{-1}) were prepared using a modified protocol[33]. The plugs were rinsed in $TE_{10:5}$ (10 mM TRIS-HCl, pH 7.8, 5 mM EDTA, pH 8.0) and then either subjected to pulsed-field gel electrophoresis, digested with restriction enzymes and end-cloned, or stored at 4 °C.

DNA labelling. DNA from YACs, P1s and cDNA clones was labelled for hybridization experiments using one of three standard techniques. For *in situ* hybridization P1 clones were biotinylated by nick-translation as described[34]. For cDNA screening the probes were radio-labelled with α-^{32}P-dCTP using random-oligo primers[35]. For P1 library screening DNA probes were labelled by directly incorporating ^{32}P-dCTP by PCR into the PCR product.

YAC analysis. Yeast chromosomes prepared in agarose plugs were size-separated on 1% SeaKem agarose gels in 0.5× TBE using a CHEF DRII Mapper (BioRad). Gels were blotted onto Biotrans Nylon membranes (ICN) or HyBond-N (Amersham) using alkaline blotting, and UV cross-linked as directed by the manufacturer. Filters were hybridized and washed according to standard procedures, then exposed to Kodak XAR film. The sizes of individual clones were determined by comparison to their relative positions among the natural yeast chromosomes.

YAC end-cloning. Inverse PCR and end-clone sequencing was performed as described[36]. Also, a new primer was developed specifically for use with *Rsa*I-digested substrate (*Rsa*-YAC3L: CAG GAA ACA GCT ATC ACC GGA AGA ACG AAG GAA GGA GC). The inverse PCR technique was adapted for the P1 vector (pAd10sacBII) using the following modifications: 100–200 ng P1 DNA was digested with either *Alu*I, *Hha*I, *Hpa*II, *Nla*III, *Rsa*I or *Taq*I, phenol-extracted, and ethanol-precipitated. Lyophilized DNA pellets were resuspended in 10 µl of $TE_{10:1}$ and the DNA was ligated at a concentration of 2 µg ml^{-1}. The ligation reactions were diluted to 0.4 µg ml^{-1} and 1 µl of each was used as the template for PCR amplification of end-clones. The P13Rup (TGT AAA ACG ACG GCC AGT GGC CGC TAA TAC GAC TCA CTA) and P13Lrp (CAG GAA ACA GCT ATC ACC GCA ATA TAG TCC TAC AAT GTC) primers were used with templates digested with *Alu*I, *Hha*I, *Hpa*II, *Nla*III or *Taq*I; P15Rrp (CAG GAA ACA GCT ATC ACC GGA TCG AAA CGG CAG ATC GCA) and P15Lup (TGT AAA ACG ACG GCC AGT TAA TTG GCC GTC GAC ATT TAG) were used when templates were digested with *Nla*III and *Rsa*I. For Alu-vector PCR the DNA from two agarose plugs was heated at 68 °C in 400 µl distilled water.

Aliquots of 1–3 μl (~5 ng) were used in PCR with one vector primer (YAK3R or YAK5L)[36] and one of the Alu-primer[37,38]. YAC DNA was also amplified using the Alu primer alone and analysed adjacent to the corresponding Alu-vector products on a 2% agarose gel. Unique Alu-vector fragments were excised from the gel and prepared for DNA sequencing.

End-clone sequencing. End-clone PCR products from either inverse PCR or alu-vector PCR were purified with Centricon-100 microconcentrators (Amicon) according to manufacturer's directions. Since the PCR primers were designed with M13 universal primer (-21M13) and M13 reverse primer (M13RP1) sequencing tails, the end-clones could be sequenced on an automated sequencer ABI 373A (Applied Biosystems) without cloning.

***In situ* hybridization.** Interphase nuclei were obtained from the cultured human skin fibroblast cell line F-6606, prepared for FISH, and hybridized as described[34] with the following modifications. The hybridization signal from P1 DNA was detected using either an FITC-anti-digoxigenin antibody (Boehringer-Mannheim) for digoxigenin-labelled probes, or streptavidin-Cy3 (Jackson Immuno Research Laboratories, Inc.) for biotin-labelled probes, and the nuclei were counter-stained with DAPI. Fluorescence was visualized using an Olympus BX50 epifluorescence microscope equipped with filter sets specific for FITC, DAPI and Cy3 (Chroma Technology). Images were collected and processed using a cooled charge-coupled device (CCD) and software specially designed for FISH analysis (BDS, Inc.). The P1 clones were labelled by nick translation with either digoxigenin 11-dUTP (Boehringer Mannheim) or biotin 14-dATP (Gibco BRL); the order of the P1 clones was determined by hybridizing two P1 clones labelled in red together with one P1 clone labelled in green, and scoring the position of the green site relative to the two red sites in 100 or more interphase nuclei. The percentage of nuclei in which the green signal was observed either between or outside the two red signals was recorded.

cDNA screening. A commercially available fetal-brain cDNA library (Stratagene, no. 936206) was screened with isolated YAC or P1 clones. Phage plating, filter lifts and clone purification were all performed according to standard procedures.

Mutation screening. To screen for mutations, we designed oligo primers with sequencing tails for each gene of interest. As a template for mutation detection, RT-cDNA was prepared from RNA isolated from lymphoblast cultures from selected patients using the RNA PCR kit from Perkin Elmer. The PCR products were gel-purified and subsequently prepared for either *Taq* cycle sequencing or T7 solid phase sequencing on the ABI373A sequencing automat. Because the RNA recovered from the cell culture is a mixture of transcripts from the two homologues, basepair differences between the homologous transcripts can be observed as two overlapping peaks.

Development of PCR-based markers. New PCR primers were developed from the DNA sequence derived from the extremities of large-insert clones using a locally developed computer algorithm, OLIGO. The localization of a new STS was verified using a panel of somatic cell-hybrids (Coriell Laboratories). An STS found to lie on chromosome 17 was further tested against 5 ng of each of the isolated YAC and P1 clones to identify regions of overlap. New STSs developed from end-clones that were not positive for other existing clones were used to screen the libraries to identify additional clones. All PCR primers and conditions are available electronically by anonymous ftp (see below).

DNA sequence comparison. To identify homologies or identities between sequence we obtained and other sequences, we used the BLAST algorithm[39] via the internet. This allowed us to compare our sequence on both the DNA and protein levels, with all sequence data stored at NCBI. Based on these results, sequences were selected and retrieved for further analysis from NBCI using the E-mail retrieve server.

Electronic information access. Detailed information about PCR markers and genomic clones is available from databases connected to the internet. Some YAC information is accessible on World Wide Web (WWW) database servers at Michigan Human Genome Center (http://mendel.hgp.umich.edu/Home.html) and Baylor College of Medicine Genome Center (http://gc.bcm.tmc.edu:8088/). DNA primer sequences and PCR conditions are available from GDB. Sequence information relating to clone extremities and cDNAs is available from GenBank. The sequence, gene and primer information may also be obtained directly from the authors via anonymous ftp to corona.med.utah.edu (128.110.231.1).

Acknowledgements

We acknowledge the CEPH for providing PCR pools for YAC library screening and a collection of chromosome 17-specific YACs, the UK Human Genome Mapping Project for assistance with screening the ICI YAC library and the Michigan Genome Center for providing YACs from the St. Louis YAC library. We are grateful to Ed Meenen for synthesizing oligonucleotide primers, Dave Fuhrman and Rob Sargent for computer support and Ruth Foltz for editing the manuscript. This work was supported by NIH grants R01-HG00367-04 and R01-HG00339. S.A.S. is supported by an EMBO fellowship and S.M. by an EC fellowship. R.W. is an Investigator of the Howard Hughes Medical Institute.

Received 4 April; accepted 2 June 1994.

1. Hall, J.M. *et al.* Linkage of early-onset familial breast cancer to chromosome 17q21. *Science* **250**, 1684–1689 (1990).
2. Easton, D.F. *et al.* Genetic linkage analysis in familial breast and ovarian cancer: results from 214 families. *Am. J. hum. Genet.* **52**, 678–701 (1993).
3. Bowcock, A.M. *et al.* THRA1 and D17S183 flank an interval of <4 cM for the breast-ovarian cancer gene (*BRCA1*) on chromosome 17q21. *Am. J. hum. Genet.* **52**, 718–722 (1993).
4. Chamberlain, J.S. *et al. BRCA1* maps proximal to D17S579 on chromosome 17q21 by genetic analysis. *Am. J. hum. Genet.* **52**, 792–798 (1993).
5. Anderson, L.A. *et al.* High-density genetic map of the *BRCA1* region of chromosome-17q12–q21. *Genomics* **17**, 618–623 (1993).
6. Albertsen, H. *et al.* Genetic mapping of the *BRCA1* region on chromosome 17q21. *Am. J. hum. Genet.* **54**, 516–525 (1994).
7. Abel, K.J. *et al.* A radiation hybrid map of the *BRCA1* region of chromosome-17q12–q21. *Genomics* **17**, 632–641 (1993).
8. O'Connell, P. *et al.* A radiation hybrid map of the *BRCA1* region. *Am. J. hum. Genet.* **54**, 526–534 (1994).
9. Fleijter, W.L. *et al.* Multicolor FISH mapping with Alu-PCR-amplified YAC clone DNA determines the order of markers in the *BRCA1* region on chromosome 17q12–q21. *Genomics* **17**, 624–631 (1993).
10. Albertsen, H.M. *et al.* Construction and characterization of a human yeast artificial chromosome library containing seven haploid genome equivalents. *Proc. natn. Acad. Sci. U.S.A.* **87**, 4256–4260 (1990).
11. Chumakov, I.M. *et al.* Isolation of chromosome 21-specific yeast artificial chromosomes from a total human genome library. *Nature Genet.* **1**, 222–225 (1992).
12. Anand, R., Riley, J.H., Smith, J.C. & Markham, A.F. A 3.5 genome equivalent multi access YAC library: construction, characterisation, screening and storage. *Nucl. Acids Res.* **18**, 1951–1955 (1990).
13. Burke, D.T., Carle, G.F. & Olson, M.V. Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **236**, 806–812 (1987).
14. Shepherd, N.S. *et al.* Preparation and screening of an arrayed human genomic library generated with the P1 cloning system. *Proc. natn. Acad. Sci. U.S.A.* **91**, 2629–2633 (1994).
15. Xu, W., Rider, S.H., Tanner, M. & Solomon, E. Mapping of human chromosome 17 using a combination of pulsed field gel electrophoresis and chromosome mediated gene transfer. *Cytogenet. cell Genet.* **51** 1111–1112 (1989).
16. Lemons, R.S. *et al.* Cloning and characterization of the t(15;17) translocation breakpoint region in acute promyelocytic leukemia. *Genes Chrom. Cancer* **2**, 79–87 (1990).
17. Simard, J. *et al.* Genetic mapping of the breast-ovarian cancer syndrome to a small interval on chromosome 17q12–21: exclusion of candidate genes EDH17B2 and RARA. *Hum. molec. Genet.* **2**, 1193–1199 (1993).
18. Goldgar, D.E. *et al.* A large kindred with 17q-linked breast and ovarian cancer: genetic, phenotypic, and genealogical analysis. *J. natn. Cancer Inst.* **86**, 200–209 (1994).
19. Campbell, I.G. *et al.* A novel gene encoding a B-box protein within the BRCA1 region at 17q21.1. *Hum. molec. Genet.* **3**, 589–594 (1994).
20. Chavrier, P., Parton, R.G., Haun, H.P., Simons, K. & Zerial, M. Localization of low molecular weight GTP binding proteins to exocytic and endocytic compartments. *Cell* **62**, 317–329 (1990).
21. Nikaido T. *et al.* Cloning and nucleotide sequence of cDNA for Ki antigen, a highly conserved nuclear protein detected with sera from patients with systemic lupus erythematosus. *Clin. exp. Immunol.* **79**, 209–214 (1990).
22. Georgakopoulos, T. & Thireos, G. Two distinct yeast transcriptional activators require the function of the GCN5 protein to promote normal levels of transcription. *EMBO J.* **11**, 4145–4152 (1992).
23. Sprinkle, T.J., Kouri, R.E., Fain, P.D., Stoming, T.A. & Whitney, J.B. Chromosomal mapping of the human CNP gene using a meiotic crossover DNA panel, PCR, and allele-specific probes. *Genomics* **16**, 542–5 (1993).
24. Tonin, P. *et al.* The human insulin-like growth factor-binding protein 4 gene maps to chromosome region 17q12–q21.1 and is close to the gene for hereditary breast/ovarian cancer. *Genomics* **18**, 414–417 (1993).
25. Tsai-Pflugfelder, M. *et al.* Cloning and sequencing of cDNA encoding human DNA topoisomerase II and localization of the gene to chromosome region 17q21–22. *Proc. natn. Acad. Sci. U.S.A.* **85**, 7177–7181 (1988).
26. Elshoubagy, N. *et al.* Cloning and expression of a human ATP-citrate lyase cDNA. *Eur. J. Biochem.* **204**, 491–499 (1992).
27. Kohn, K. *et al.* Cloning and structural determination of human peptide YY cDNA and gene. *Biochem. Biophys. Acta* **1173**, 345–349 (1993).
28. Kelsell, D.P., Black, D.M., Bishop, D.T. & Spurr, N.K. Genetic analysis of the *BRCA1* region in a large breast/ovarian family: refinement of the minimal region containing *BRCA1*. *Hum. molec. Genet.* **2**, 1823–1828 (1993).
29. Smith, S.A. *et al.* Localization of the breast-ovarian cancer susceptibility gene (*BRCA1*) on 17q12–21 to an interval of ≤ 1 cM. *Genes Chrom. Cancer.* **10**, 71–76 (1994).
30. Smith, S.A., Easton, D.F., Evans D.G.R. & Ponder, B.A.J. Allele losses in the region 17q12–21 in familial breast and ovarian cancer involve the wild-type chromosome. *Nature Genet.* **2**, 128–131. (1992).
31. Cropp, C.S. *et al.* Evidence for involvement of *BRCA1* in sporadic breast carcinomas. *Cancer Res.* **54**, 2548–2551 (1994).
32. Green, E.D. & Olson, M.V. Systematic screening of yeast artificial-chromosome libraries by use of the polymerase chain reaction. *Proc. natn. Acad. Sci. U.S.A.* **87**, 1213–1217 (1990).
33. Riley, J. *et al.* A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. *Nucl. Acids Res.* **18**, 2887–2890 (1990).
34. Trask, B.J., Massa, H., Kenwick, S. & Gitschier, J. Mapping of human chromosome Xq28 by two-color fluorescence *in situ* hybridization of DNA sequences to interphase cell nuclei. *Am. J. hum. Genet.* **48**, 1–15 (1991).
35. Feinberg, A.P. & Vogelstein, B. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132**, 6–13 (1984).
36. Joslyn, G. *et al.* Identification of deletion mutations and three new genes at the familial polyposis locus. *Cell* **66**, 601–613 (1991).
37. Nelson, D.L. *et al.* Alu polymerase chain reaction: a method for rapid isolation of human-specific sequences from complex DNA sources. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6686–6690 (1989).
38. Tagle, D.A. & Collins, F.S. An optimized Alu-PCR primer pair for human-specific amplification of YACs and somatic cell hybrids. *Hum. molec. Genet.* **1**, 121–122 (1992).
39. Altschul, S.F., Gish, W., Miller, W., Myers, E.W. & Lipman, D.J. Basic local alignment search tool. *J. molec. Biol.* **215**, 403–410 (1990).
40. Emi, M. *et al.* A novel metalloprotease/disintegenn-like gene at 17q21.3 is somatically rearranged in two primary breast cancers. *Nature Genet.* **5**, 151–157 (1993).

GENOMICS 33, 207–213 (1996)
ARTICLE NO. 0185

Sequence, Genomic Structure, and Chromosomal Assignment of Human DOC-2

HANS M. ALBERTSEN,*,†,[1] SIMON A. SMITH,† ROBERTA MELIS,† BRIANA WILLIAMS,* PILAR HOLIK,† JEFF STEVENS,† AND RAY WHITE*,†

*Eccles Institute of Human Genetics and †Huntsman Cancer Institute, University of Utah, Salt Lake City, Utah 84112

Received August 21, 1995; accepted December 21, 1995

DOC-2 is a human gene originally identified as a 767-bp cDNA fragment isolated from normal ovarian epithelial cells by differential display against ovarian carcinoma cells. We have now determined the complete cDNA sequence of the 3.2-kb DOC-2 transcript and localized the gene to chromosome 5. A 12.5-kb genomic fragment at the 5′-end of DOC-2 has also been sequenced, revealing the intron–exon structure of the first eight exons (788 bases) of the DOC-2 gene. Translation of the DOC-2 cDNA predicts a hydrophobic protein of 770 amino acid residues with a molecular weight of 82.5 kDa. Comparison of the DNA and amino acid sequences of DOC-2 to publicly accessible sequence databases revealed 83% identity to p96, a murine protein of similar size, thought to be a mitogen-responsive phosphoprotein. In addition, about 45% identity was observed between the first 140 N-terminal residues of DOC-2 and the *Caenorhabditas elegans* M110.5 and *Drosophila melanogaster* Dab genes.

INTRODUCTION

Genes that show differential expression between normal and tumor tissues are likely to function either directly in growth regulation or cellular differentiation or indirectly as a response to changes in the cellular environment. Those genes that directly influence growth or differentiation can be grouped into two separate classes, the first type functioning as negative regulators (the prominent group known as tumor suppressor genes) and the second functioning to stimulate growth and differentiation. Several of the tumor suppressor genes characterized to date, including *RB* (Lee *et al.,* 1987), *APC* (Groden *et al.,* 1991), and *BRCA1* (Miki *et al.,* 1994), control growth in epithelial cells; accumulated evidence suggests that additional growth suppressors will be identified in various epithelial cell types. One candidate for such a role, DOC-2, was identified by means of the differential display technique (Liang and Pardee, 1992), and DOC-2 was shown to be expressed in all normal ovarian epithelial cells but significantly down-regulated or absent in all of a series of ovarian carcinoma cell lines tested (Mok *et al.,* 1994).

The biological function of DOC-2 remains unclear, but the predicted protein sequence shows some degree of homology to that of proteins from other species. At the amino-terminal end of DOC-2, a 140 amino acid segment shares homology with a recently described phosphotyrosine interaction domain (PID) (Bork and Margolis, 1995). Recently published results indicate that the mouse homologue of DOC-2, p96, is phosphorylated on serine residues (but not on tyrosine residues) in a pattern that appears to be linked to the cell cycle: a minimal degree of phosphorylation occurs in the G_1 stage and rapidly increases following mitogenic stimulation with CSF-1 (Xu *et al.,* 1995).

Originally, only a 767-bp fragment of the DOC-2 cDNA was identified. Here, we present the complete ~3200-bp sequence of DOC-2, its chromosomal location, and a 12.3-kb contiguous genomic sequence harboring the first eight exons coding for the N-terminal third of the DOC-2 protein. We also demonstrate that DOC-2 is expressed in a wide variety of tissues.

MATERIALS AND METHODS

Isolation of human DOC-2 cDNA clones. cDNA clones were isolated from two commercially available cDNA Lambda ZAP phage libraries: a brain-stem library and a fetal-retina library (Stratagene, San Diego, CA, Catalog Nos. 936206 and 937202). Each library was plated at a density of approximately 25,000 pfu per 150-mm petri dish on *Escherichia coli* strain LE392. Phage plaques were lifted on Biotrans nylon membranes (ICN) or Hybond-N (Amersham) and screened according to standard procedures. DNA probes for hybridization were radioactively labeled with [α-^{32}P]dCTP using a Prime-It II DNA labeling kit (Stratagene, Catalog No. 300385). Filters were washed and exposed to HyperFilm (Amersham) overnight at −70°C. Plasmids were excised from the phage according to the manufacturer's recommendations.

DNA sequencing. Bluescript plasmids or PCR products designed with M13 universal primer (−21M13) and M13 reverse primer

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under Accession Nos. U39050, U41096, and U41111.

[1] To whom correspondence should be addressed. Telephone: (801) 585-6178. Fax: (801) 585-3833. E-mail: Hans.Albertsen@genetics.utah.edu.

0888-7543/96 $18.00

(M13RP1) sequencing tails were sequenced using the dye-primer technique on an ABI 373A automated sequencer (Applied Biosystems). Alternatively, plasmids were sequenced on the ABI 373A instrument using a dye-terminator technique.

DNA and protein sequence analysis. DNA sequences were merged and edited using the Intelligenetics suite of programs. Searches for protein alignments and motifs were performed using the Wisconsin Sequence Analysis Package (Version 8). To identify homologies or identities between the DOC-2 sequence and other sequences, we used the BLAST algorithm (Altschul *et al.,* 1990) through the Internet. This procedure allowed for sequence comparisons on both the DNA and protein levels, against all sequences stored at NCBI. On the basis of these results, we retrieved selected sequences from NCBI for further analysis using the E-mail retrieve server.

RNA preparation and first-strand cDNA synthesis. RNA was prepared from three different solid tissues as well as cultured cells from two epithelial cell lines, using the TRIzol reagent (Gibco BRL, Catalog No. 15596-026) according to a protocol provided by the manufacturer. First-strand cDNAs were synthesized from a RNA template using the SuperScript II RT kit (Gibco BRL, Catalog No. 18064-014).

DNA primers, primer design, and PCR amplification. DNA primers were designed to be 18–21 bases long with calculated melting points between 56 and 60°C. Tailed sequencing primers deviated from this pattern by additionally having the 18-base-long universal forward or reverse primer sequences attached at the 5′-end of each locus-specific primer. Bases from the sequencing tails were ignored for the purpose of determining PCR annealing temperatures. The oligonucleotide primers were synthesized on an ABI394 DNA/RNA synthesizer (Applied Biosystems), lyophilized, and resuspended in H_2O to a concentration of 20 μM. Primers used for direct sequencing by means of the chain terminator technique were purified on OPC columns (Applied Biosystems) following the manufacturer's recommendations. In general, PCR amplification was performed in a Perkin–Elmer 9600 ThermoCycler in a PCR buffer with $MgCl_2$ to a final concentration of 1.5 mM, under these standard thermocycling conditions:

Initial denaturation:		94°C for	120 s
30×	Denaturation:	94°C for	20 s
	Annealing:	Lower T_m of the two primers for	20 s
	Extension:	72°C for	40 s

PCR products were visualized in standard 150-ml agarose gels containing 1× TBE, 0.8–2.0% SeaKem (FMC BioProducts), and 1 μl ethidium bromide.

In situ hybridization. Metaphase chromosomes were prepared from peripheral blood from a normal male and hybridized with the P1 clone, 35G12, using the general procedures described by Lichter *et al.* (1991). The hybridization signal from the biotinylated P1 DNA was detected using streptavidin-Cy3 (Jackson Immuno Research Laboratories, Inc.), and the chromosomes were counterstained with DAPI (4′-6-diamidino-2-phenylindole). Fluorescence was visualized using an Olympus BX50 epifluorescence microscope equipped with filter sets specific for DAPI and Cy3 (Chroma Technology). Images were collected and processed using a cooled charge-coupled device and software specially designed for FISH analysis (Vysis, Inc.).

Somatic cell hybrid analysis. The NIGMS human/rodent somatic cell hybrid mapping panel 2 and the regional mapping panel for chromosome 5 (Coriell Cell Repository) were tested with the DNA oligo primer pair doc2-B (AAATTTTGGAGAGTCTAGAGC) and doc2-C (GAATACGCTTGGTTCGTCC). Each reaction contained 50 ng of template DNA and 12.5 pmol of each primer in a 25-μl reaction volume. The PCR buffer contained $MgCl_2$ to a final concentration of 2.0 mM, and the amplification was carried out under the following thermocycling conditions:

Initial denaturation:		94°C for	120 s
30×	Denaturation:	94°C for	30 s
	Annealing:	52°C for	30 s
	Extension:	72°C for	40 s
Final extension:		72°C for	120 s

RESULTS AND DISCUSSION

Isolation of cDNA Clones from the DOC-2 Locus

During a search for expressed genes in the BRCA1 region, several novel transcripts were isolated using genomic clones to screen cDNA libraries (Albertsen *et al.,* 1994). One of the cDNA clones isolated in this manner, 40F1, was found to be chimeric; 542 bp at one end of the insert derived from the DOC-2 gene, and the remaining 1888 bp derived from the DLG3 gene (Smith *et al.,* 1995). The predicted protein translation of the chimeric clone 40F1 showed a single open reading frame that included the correct orientations and frames from both genes. Rescreening of the cDNA libraries with 40F1 resulted in the isolation of several new cDNA clones, most of which were transcripts of the DLG3 gene. However, a 2726-bp clone, 1RA1, which corresponded uniquely to the DOC-2 gene, was also isolated.

Sequence Analysis of 1RA1

The complete sequence of 1RA1 was determined by first sequencing the two extremities of the clone using the standard universal forward and reverse sequencing primers; then, using custom-made primers, the internal sequence of 1RA1 was completed following several rounds of sequence walking. The consensus cDNA sequence derived from combining 1RA1 and the originally published partial sequence (GenBank Accession No. L16886) extended for 2960 bp, with the L16886 DNA sequence extending 241 bases further 5′ than 1RA1. Analysis of this sequence revealed an open reading frame extending from the extreme 5′-end of the sequence for 2163 bp, followed by 792 bp of 3′-untranslated sequence and a poly(A) tail. Of four candidate initiation methionine codons located at the 5′-end of the cDNA, none were in the consensus environment described for translational initiation (Kozak, 1991), nor were any in-frame stop codons present in the 5′-end of the predicted reading frame. This suggested that the 5′-end of the DOC-2 cDNA remained to be found.

Isolation and Sequencing of Genomic Clones Containing the 5′-End of DOC-2

To isolate genomic clones containing the 5′-end of DOC-2, the 1RA1 cDNA was used to screen a P1 phage library (Shepherd *et al.,* 1994). Two clones, 35G12 and 47E3, were obtained, from which restriction fragments generated by the enzymes *Eco*RI, *Hin*dIII, and *Pst*I were cloned into pBluescriptII. Four subclones derived from 47E3 were selected by hybridization with a radiolabeled fragment from the 5′-end of the DOC-2 cDNA and sequenced as described above. The sequences we obtained fell into two contigs of 2073 and 12518 bp respectively.

Identification of the 5′-End of DOC-2 cDNA and Detection of Eight Intron–Exon Boundaries

Since attempts at isolating other cDNA clones that could extend the DOC-2 cDNA sequence further 5′ were unsuccessful, we sought to identify the 5′-end of DOC-2 by comparing the genomic DNA sequence with that of murine p96 (GenBank Accession No. U18869), which we believe is the homologue of DOC-2 (see below). This analysis revealed a single region in the human genomic sequence that was highly homologous to the DNA sequence of the first exon of p96. To verify that the homologous sequence indeed corresponded to the 5′-end of the DOC-2 cDNA, we developed four PCR primers (doc2-5A, GTCTAATACAGCAGGAGAAGG; -5B, CTAAACTATGCCTACAGGTGTC; -5C, GGGATCGCCTGGTGTCACCAA; and -5D, GTCCACTGGTACTGAGGTTTG) located in the genomic sequence of DOC-2, upstream of the predicted translational initiation site. The primers were sequentially positioned, with doc2-5A being farthest away from the predicted translational initiation site. Each of the PCR primers was tested separately with a second primer (40Irp, TTGTCCCTGAGACCGACCA) located in the third exon of DOC-2, using first-strand cDNA prepared from small intestine, arm skin, fallopian tube, PPC1 (a prostate cancer cell line), and SF15-2 (a derivative of PPC1 exhibiting reduced malignant growth characteristics as a result of introduction of normal human chromosome 17q (Murakami *et al.,* 1995)) as template. No products were observed with doc2-5A, while doc2-5B showed amplification using template prepared from fallopian tube only. Both doc2-5C and doc2-5D gave rise to PCR products from each of the templates tested (data not shown). Sequence analysis of several of these PCR products confirmed the high degree of homology between the 5′-end of DOC-2 and that of p96 and extended the DOC-2 ORF by an additional 146 bp (the locations of doc2-5B, -5C, -5D, and 40Irp are indicated in Fig. 1). The full-length DOC-2 cDNA sequence, along with the genomic sequences from the 5′-end of the DOC-2 gene, are available from GenBank under Accession No. U39050. Further sequence comparison between the cDNA sequence and the genomic sequence of DOC-2 enabled us to determine the genomic structure of the amino terminal third of the gene, including the exact positions of the first eight intron–exon boundaries (see Fig. 1).

Homology of DOC-2 with Murine p96 and Discovery of Alternative Splice Forms

Comparison of the predicted amino acid sequence of the DOC-2 protein with sequences present in the NCBI database revealed very high homology with murine p96 (U18869). The homology with p96 was strongest at the amino terminal of DOC-2 corresponding to the phosphotyrosine domain (Bork and Margolis, 1995), but appeared to weaken in the central portion and toward the carboxy terminal of the protein. The region between basepairs 1503 and 1594 in the human sequence appeared to be shifted by a single nucleotide, resulting in a local shift of reading frame between DOC-2 and p96. To verify the sequence of the p96 gene, total RNA extracted from a differentiated and an undifferentiated mouse stem cell line was reverse-transcribed and used as template in a RT-PCR reaction with a pair of primers specific to the mouse (the RNA was generously donated by Dr. Suzi Mansour). The DNA sequence of the resulting PCR product indicated that two errors had been made during the original sequencing of p96. This result has been confirmed independently, and the DNA sequence of p96 has been corrected (Dr. Charles O. Rock, St. Jude Children's Research Hospital, Memphis TN, pers. comm., Oct. 24, 1995). Translation of the DOC-2 cDNA predicts a hydrophobic protein of 770 amino acid residues with a molecular weight of 82.5 kDa sharing 81% identity with the gene product of p96, which is predicted to encode a protein of 766 amino acid residues with a molecular weight of 82.7 kDa.

Only a single methionine is present in the first exon of DOC-2, and its location is identical to the predicted translational start site of p96. Only 1 of the 31 amino acids encoded by the first exon of DOC-2 varies with the mouse sequence. However, the DNA sequences immediately upstream of the translational start sites show extensive divergence. Also, the sequences relevant to translational initiation show some degree of variation between the two species, and each conforms only moderately well to the consensus sequence proposed by Kozak (1991). Clone 1RA1 has a poly(A) tail at its 3′-end, but the presumptive polyadenylation signal approximately 15 bases upstream of this poly(A) tail (at position 3185) is somewhat degenerate (AACAGA as opposed to AAUAAA), although it retains the correct composition of five purines and one pyrimidine. The possibility that this is not an artifact of 1RA1 is supported by several expressed sequence tags (EST) (e.g., R37515) that share the same polyadenylation site. However, investigation of several other retrieved EST sequences (e.g., R63200) indicates that an additional polyadenylation site (at positions 2787) may also be used (see Fig. 1).

Evidence indicates that at least one variant of DOC-2 mRNA exists, probably as an alternative splice form. When we compared the retinal cDNA clone, 1RA1, to the ovarian clone (L16886) identified by differential display, we noted the absence of a 60-bp exon between bases 729 and 789 in the cDNA sequence. Comparison of this presumed splice variant with the databases revealed that the same splice variant of the mouse p96 gene exists, under the name p93, almost certainly confirming the variant as an alternative splice form (see Fig. 1).

A second sequence variation was also detected as a

```
GTC TGT GGG AGG TTA TGT TTA TTT GAG ACT TCT CCA TCG GGA TCG CCT GGT GTC ACC AAG TGT CCA CTG GTA CTG AGG TTT GCT GCC TGC CTT CTT GCC ATG TCT AAC GAA GTA GAA ACA   120
doc2-5B                                        doc2-5C                        doc2-5D                                  MET Ser Asn Glu Val Glu Thr

AGT GCA ACC AAT GGT CAG CCC GAC CAA CAG GCC GCA CCA AAA GCA CCC TCA AAG AAG GAA AAA AAG AAA GGC CCT GAA AAG ACA GAT GAA TAT CTC TTA GCA AGG TTC AAA GGC GAT GGT   240
Ser Ala Thr Asn Gly Gln Pro Asp Gln Gln Ala Ala Pro Lys Ala Pro Ser Lys Lys Glu Lys Lys Lys Gly Pro Glu Lys Thr Asp Glu Tyr Leu Leu Ala Arg Phe Lys Gly Asp Gly

GTA AAA TAT AAG GCC AAG CTG ATT GGC ATT GAT GAT GTG CCA GAT GCA AGA GGG GAT AAA ATG AGC CAA GAC TCT ATG ATG AAA CTA AAG GGA ATG GCG GCA CGT GGT CGG TCT CAG GGA   40Irp
Val Lys Tyr Lys Ala Lys Leu Ile Gly Ile Asp Asp Val Pro Asp Ala Arg Gly Asp Lys Met Ser Gln Asp Ser Met Met Lys Leu Lys Gly Met Ala Ala Arg Gly Arg Ser Gln Gly

CAA CAC AAA CAA AGG ATC TGG GTC AAC ATT TCC CTT TCT GGG ATA AAA ATA ATT GAT GAG AAA ACT GGG GTA ATA GAG CAT GAA CAT CCA GTA AAT AAG ATT TCT TTC ATT GCC CGT GAT   480
Gln His Lys Gln Arg Ile Trp Val Asn Ile Ser Leu Ser Gly Ile Lys Ile Ile Asp Glu Lys Thr Gly Val Ile Glu His Glu His Pro Val Asn Lys Ile Ser Phe Ile Ala Arg Asp

GTG ACA GAC AAC CGG GCA TTT GGT TAC GTG TGT GGA GGA GAA GGC CAG CAT CAG TTT TTT GCC ATA AAA ACC GGG CAA CAG GCT GAA CCA TTA GTT GTT GAT CTT AAA GAC CTT TTT CAA   600
Val Thr Asp Asn Arg Ala Phe Gly Tyr Val Cys Gly Gly Glu Gly Gln His Gln Phe Phe Ala Ile Lys Thr Gly Gln Gln Ala Glu Pro Leu Val Val Asp Leu Lys Asp Leu Phe Gln

GTT ATC TAT AAT GTA AAG AAA AAG GAA GAA GAA AAG AAA AAG ATA GAG GAA GCC AGC AAA GCA GTT GAG AAT GGG AGT GAG GCC CTA ATG ATT CTA GAT GAC CAA ACT AAC AAA CTG AAA   720
Val Ile Tyr Asn Val Lys Lys Lys Glu Glu Glu Lys Lys Lys Ile Glu Glu Ala Ser Lys Ala Val Glu Asn Gly Ser Glu Ala Leu Met Ile Leu Asp Asp Gln Thr Asn Lys Leu Lys

TCG GGT GTT GAC CAG ATG GAT TTG TTT GGG GAC ATG TCT ACA CCT CCT GAC CTA AAT AGT CCA ACA GAA AGC AAA GAT ATC CTG TTA GTG GAT CTA AAC TCT GAA ATC GAC ACC AAT CAG   840
Ser Gly Val Asp Gln Met Asp Leu Phe Gly Asp Met Ser Thr Pro Pro Asp Leu Asn Ser Pro Thr Glu Ser Lys Asp Ile Leu Leu Val Asp Leu Asn Ser Glu Ile Asp Thr Asn Gln

AAT TCT TTA AGA GAA AAT CCA TTC TTA ACA AAC GGC ATC ACC TCC TGT TCT CTT CCT CGA CCA ACG CCT CAG GCA TCC TTC TCG CCT GAA AAT GCC TTT TCT GCC AAT CTC AAC TTC TTT   960
Asn Ser Leu Arg Glu Asn Pro Phe Leu Thr Asn Gly Ile Thr Ser Cys Ser Leu Pro Arg Pro Thr Pro Gln Ala Ser Phe Ser Pro Glu Asn Ala Phe Ser Ala Asn Leu Asn Phe Phe

CCC ACC CCT AAT CCT GAT CCT TTC CGT GAC GAT CCT TTC ACA CAG CCA GAC CAA TCG ACA CCT TCT TCG TTT GAT TCT CTC AAA TCT CCA GAT CAG AAG AAA GAG AAT TCG AGT AGC TCG  1080
Pro Thr Pro Asn Pro Asp Pro Phe Arg Asp Asp Pro Phe Thr Gln Pro Asp Gln Ser Thr Pro Ser Ser Phe Asp Ser Leu Lys Ser Pro Asp Gln Lys Lys Glu Asn Ser Ser Ser Ser

TCT ACT CCG CTG AGT AAT GGG CCC CTG AAT GGT GAT GTT GAC TAC TTT GGT CAG CAA TTT GAC CAG ATC TCT AAC CGG ACT GGC AAA CAG GAA GCT CAG GCA GGC CCA TGG CCC TTT TCA  1200
Ser Thr Pro Leu Ser Asn Gly Pro Leu Asn Gly Asp Val Asp Tyr Phe Gly Gln Gln Phe Asp Gln Ile Ser Asn Arg Thr Gly Lys Gln Glu Ala Gln Ala Gly Pro Trp Pro Phe Ser

AGT TCG CAA ACC CAG CCA GCA GTG AGA ACT CAA AAT GGG GTA TCT GAA AGA GAA CAG AAC GGC TTC TCT GTC AAA TCC TCC CCG AAC CCT TTT GTG GGA AGC CCT CCC AAA GGA CTG TCC  1320
Ser Ser Gln Thr Gln Pro Ala Val Arg Thr Gln Asn Gly Val Ser Glu Arg Glu Gln Asn Gly Phe Ser Val Lys Ser Ser Pro Asn Pro Phe Val Gly Ser Pro Pro Lys Gly Leu Ser

ATA CAG AAT GGC GTA AAG CAG GAC TTG GAA AGC TCT GTC CAG TCC TCA CCA CAT GAC TCC ATA GCC ATT ATC CCA CCT CCA CAA AGT ACC AAA CCA GGA AGA GGC AGA AGG ACT GCT AAG  1440
Ile Gln Asn Gly Val Lys Gln Asp Leu Glu Ser Ser Val Gln Ser Ser Pro His Asp Ser Ile Ala Ile Ile Pro Pro Pro Gln Ser Thr Lys Pro Gly Arg Gly Arg Arg Thr Ala Lys

TCT TCA GCC AAT GAC TTG CTT GCA TCA GAC ATC TTT GCT CCT CCC GTC TCA GAA CCT TCA GGC CAG GCG TCA CCC ACA GGA CAA CCT ACA GCC CTG CAG CCC AAC CCT CTG GAT CTC TTC  1560
Ser Ser Ala Asn Asp Leu Leu Ala Ser Asp Ile Phe Ala Pro Pro Val Ser Glu Pro Ser Gly Gln Ala Ser Pro Thr Gly Gln Pro Thr Ala Leu Gln Pro Asn Pro Leu Asp Leu Phe

AAA ACA AGT GCT CCT GCC CCA GTG GGG CCC CTG GTG GGT CTA GGT GGT GTA ACT GTC ACA CTC CCT CAG GCA GGA CCA TGG AAC ACA GCA TCT TTG GTC TTC AAT CAG TCC CCT TCA ATG  1680
Lys Thr Ser Ala Pro Ala Pro Val Gly Pro Leu Val Gly Leu Gly Gly Val Thr Val Thr Leu Pro Gln Ala Gly Pro Trp Asn Thr Ala Ser Leu Val Phe Asn Gln Ser Pro Ser Met

GCT CCG GGA GCC ATG ATG GGT GGT CAA CCT TCA GGT TTT AGT CAG CCC GTC ATT TTT GGT ACA AGT CCA GCT GTT TCA GGT TGG AAC CAG CCT TCA CCC TTT GCA GCC TCA ACT CCC CCT  1800
Ala Pro Gly Ala Met Met Gly Gly Gln Pro Ser Gly Phe Ser Gln Pro Val Ile Phe Gly Thr Ser Pro Ala Val Ser Gly Trp Asn Gln Pro Ser Pro Phe Ala Ala Ser Thr Pro Pro

CCA GTG CCT GTT GTC TGG GGG CCT TCT GCA TCT GTG GCA CCC AAT GCT TGG TCA ACA ACA AGC CCT TTG GGG AAT CCT TTT CAG AGC AAT ATT TTT CCA GCT CCT GCT GTG TCC ACT CAG  1920
Pro Val Pro Val Val Trp Gly Pro Ser Ala Ser Val Ala Pro Asn Ala Trp Ser Thr Thr Ser Pro Leu Gly Asn Pro Phe Gln Ser Asn Ile Phe Pro Ala Pro Ala Val Ser Thr Gln

CCC CCA TCC ATG CAC TCC TCT CTC CTG GTC ACT CCT CCT CAG CCA CCT CCC AGA GCT GGC CCT CCC AAG GAC ATC TCC AGT GAT GCC TTC ACT GCC TTA GAC CCA CTT GGG GAT AAA GAG  2040
Pro Pro Ser Met His Ser Ser Leu Leu Val Thr Pro Pro Gln Pro Pro Pro Arg Ala Gly Pro Pro Lys Asp Ile Ser Ser Asp Ala Phe Thr Ala Leu Asp Pro Leu Gly Asp Lys Glu

ATC AAG GAT GTG AAA GAA ATG TTT AAG GAT TTC CAA CTG CGG CAG CCA CCT GCT GTG CCC GCG CGG AAG GGA GAG CAG ACT TCT TCT GGG ACT TTG AGT GCC TTT GCC AGT TAT TTC AAC  2160
Ile Lys Asp Val Lys Glu Met Phe Lys Asp Phe Gln Leu Arg Gln Pro Pro Ala Val Pro Ala Arg Lys Gly Glu Gln Thr Ser Ser Gly Thr Leu Ser Ala Phe Ala Ser Tyr Phe Asn

AGC AAG GTT GGC ATT CCT CAG GAG AAT GCA GAC CAT GAT GAC TTT GAT GCT AAT CAA CTA TTG AAC AAG ATC AAT GAA CCA CCA AAG CCA GCT CCC AGA CAA GTT TCC CTG CCA GTT ACC  2280
Ser Lys Val Gly Ile Pro Gln Glu Asn Ala Asp His Asp Asp Phe Asp Ala Asn Gln Leu Leu Asn Lys Ile Asn Glu Pro Pro Lys Pro Ala Pro Arg Gln Val Ser Leu Pro Val Thr

AAA TCT ACT GAC AAT GCA TTT GAG AAC CCT TTC TTT AAA GAT TCT TTT GGT TCA TCA CAA GCC TCT GTG GCT TCT TCT CAA CCT GTA TCT TCT GAG ATG TAT AGG GAT CCA TTT GGA AAT  2400
Lys Ser Thr Asp Asn Ala Phe Glu Asn Pro Phe Phe Lys Asp Ser Phe Gly Ser Ser Gln Ala Ser Val Ala Ser Ser Gln Pro Val Ser Ser Glu Met Tyr Arg Asp Pro Phe Gly Asn

CCT TTT GCC TAA ATT CTG AAC TTG GTC TGC AGA CCA TCC AGA GGA ATA AAA AGG TTG GCC TTA GTA GTC AAA AAC AAA GCT GAT AGC CAG ACA CGT TCT GAT TTC TGC CCT TGT TCC AGC  2520
Pro Phe Ala STOP

TTT GAC GTA TTA TCT GTT GCC TTA TTT CTC ATT GCC TCT TCT ACT TGT AAA ATG CTT TTC ACT TTC TGT CTA GGT TAA AGC TAA ACT GAA TCT ATG GCT TTA AAT AAA TTA AGA TCC TAA  2640

ACT CTC TAG CTT AAG TGT AAA TGA AGT ACA GTA GTT TCC CTA CTG AAC CCT ACC TCT TGT GTC CCT GGA ACC TTC TAG AAC ACC TGC CTT CTA CCC TCT GGT TGG GAG ATG CAG CCA CCA  2760

CAT CCC TTC ATA TCA TAC TGT TTT GAA TAA ATT TTC AAA TCC TTA TTG TTC AGA GTT GTT TGG GGG TTC TGT TTC AGA GCA TAA AAC CTA AAG GTT ATA GTA GAA CAA GGC ACC TTC TTA  2880

AAA GAA ATC TTG CTT CAG ACC ATC AGT TAC AGA GAA TTT CCT AAA GTA AAA TTG AAG CAA CTA CAA CTT CTC CTT AGA CAC TTT GGA ATC TAA CCA CTT AAG GAC CTT TTT AAA GAG ATA  3000

GCT TCT CTT CTT TCT GAA GAT CAA TTT CTC CCA AGG CCA AGA TTG TCC TTT TCT CCC ATT TCT TGC TAG CTA TTG CAA ATG AGG GAA GAA CAT TAT TCA TCT CTC CTC CCC TTT TTT TTC  3120

TGA TTC TTT TTT CAG TCA GTT TTG CTC CTG GGT TCA AGT AGT ATT ACC ACC CTT TCA CAA GCA ACA GAC TCT CAC AGG GCA AAA AAA AAA AAA
```

```
          1                                                                                                  100
M110.5    MEMKTGRSCA PKGADHDFTL LSHSIPHHHL TILFGLTFVP KSFVILLVFF YCSLETMAQK SDISVETANA TSGKPNPPSP KSRLAMLKRT KKASNASSDP
Dab       .......... .......... .......... ........MV KSLV...... .AKLST..AS SNLSLA...S TFGGGSGAAE ETNYAKHR.. .......NDP
DOC-2     .......... .......... .......... .......... .......... .........M SN.EVET.SA TNGQPDQQAA PK..APSKKE KKKGPEKTDE

          101                                                                                                200
M110.5    F...RFQNNG ISYKGKLIGE QDVDKARGDA MCAEAMRTAK SII...KAAG AHKTR..ITL QINIDGIKVL DEKSGAVLHN FPVSRISFIA RDSSDARAFG
Dab       G...RFFGDG VQFKAKLIGI LEV..ARPEV IGC.ARRRCK ISK...WHPG GWRAQAAITI HVTIDGLRLR DEKTGDSLYH HPVHKISFIA QDMTDSRAFG
DOC-2     YLLARFKGDG VKYKAKLIGI DDVPDARGDK MSQDSMMKLK GMAARGRSQG QHKQR..IWV NISLSGIKII DEKTGVIEHE HPVNKISFIA RDVTDNRAFG

          201                                                                                                300
M110.5    LVYGEPGGKY KFYGIKTAQA ADQAVLAIRD MFQVVFEMKK KQIEQVKQQQ IQDGG..... ...AEI.... .SSKKEGGVA VADLLDLESE LQQIERG...
Dab       YIFGSPDSGH RFFGIKTDKA ASQVVLAMRD LFQVVFELKK KEIE.MARQQ IQGKSLHDHS SQLASL.... .SSLKSSGLG GMGL..... .........
DOC-2     YVCGG.EGQH QFFAIKTGQQ AEPLVVDLKD LFQVIYNVKK KEEE...KKK IEEASKAVEN GSEALMILDD QTNKLKSGVD QMDLFGDMST PPDLNSPTES
```

FIG. 2. Homologies between the 140 N-terminal amino acids of DOC-2 and genes from two widely divergent species, *D. melanogaster* (Dab gene) and *C. elegans* (M110.5 gene). Identities, highlighted in boldface, were found using the computer algorithm PILEUP with the parameters GapWeight set to 2.0 and GapLengthWeight set to 0.1.

result of the database searches. In this case we found three ESTs (Accession Nos. R80479, R81944, R63199) that differed from 1RA1 in the section between basepairs 2580 and 2598 in the 3′-untranslated region of DOC-2. The differences were in all cases consistent with an 18-bp inversion. We found a 5-bp palindrome flanking the inversion on both sides, which could be responsible for the inversion (see Fig. 1). We also found evidence for a 70-bp duplication in the genomic sequence during the search for exon boundaries. This duplication, which includes part of the alternatively spliced exon 8, spans the region between bases 10890 and 10942 and was inserted in the same orientation between bases 10693 and 10745, approximately 160 bases upstream of its original location (data not shown). Eleven percent variation between the duplicated sequences (8 of 70 nucleotides) suggests that the event is quite ancient.

DOC-2 Protein Homology to Other Genes

Two other proteins in the sequence databases possessed significant homology to the amino terminal of DOC-2. *Caenorhabditis elegans* gene M110.5 (Z49968) showed the highest degree of identity, with 47% in the presumptive PID, followed by the *Drosophila melanogaster* disabled (Dab) gene (L08845), with 43% identity in the same region. Figure 2 shows the result of the multisequence alignments of the amino-terminal ends of M110.5, Dab, and DOC-2.

DOC-2 Tissue Specificity

In addition to the three normal tissues and the two prostate cancer cell lines where we detected transcription of the DOC-2 mRNA using RT-PCR, DOC-2 mRNA is known to be present in normal ovarian epithelial tissue (Mok *et al.*, 1994). We also can deduce that DOC-2 must be expressed in cells from brain stem and fetal retina, since we isolated cDNA clones from libraries made from these tissues. We have thus determined that DOC-2 is expressed in at least seven different human tissues. In mouse, p96 expression has been found in differentiated and an undifferentiated mouse stem cell line (this study) and in mouse macrophage and brain cells (Xu *et al.*, 1995). Combined, these observations give the impression that DOC-2 and its murine homologue, p96, are expressed in a tissue-independent manner.

Chromosomal Mapping of DOC-2

We were able to determine the genomic location of DOC-2 to chromosome 5p13 by fluorescence *in situ* hybridization using the P1 clone 35G12 as hybridization probe against metaphase chromosomes from a normal male (see Fig. 3). The map location of DOC-2 was independently confirmed by analyzing Coriell human/rodent somatic cell hybrid mapping panel 2 in conjunction with the Coriell regional mapping panel of chromosome 5 (NIGMS, Camden, NJ). Using the oligo primers doc2-B and doc2-C, only cell lines containing the human chromosome 5p13 region provided PCR-positive template (data not shown). No amplification was expected from the rodent homologue, since the primers used in the experiment were located in intronic sequence.

CONCLUSION

In summary, the full-length sequence of the human DOC-2 gene has been ascertained, together with its

FIG. 1. Nucleotide sequence and predicted protein translation of the human DOC-2 gene. The first eight exon boundaries are indicated with vertical arrows. The location of the oligo primers (doc2-5B (only the last three bases), doc2-5C, doc2-5D, and 40Irp) used to verify the 5′-end DNA sequence of DOC-2 are shown with bold underlines of the DNA sequence. Two other regions in the translated part of the gene are underlined: the location of the 140 amino acid segment with the phosphotyrosine interaction domain (PID) (solid line under the amino acid sequence), and the location of the alternatively spliced form of DOC-2 (dotted line under the amino acid sequence). Three underlined sequences in the 3′-untranslated region of DOC-2 indicate the region of inversion and the two functional polyadenylation sites. The smallest possible inversion is indicated in boldface with the flanking 5-bp palindromes indicated with the underline.

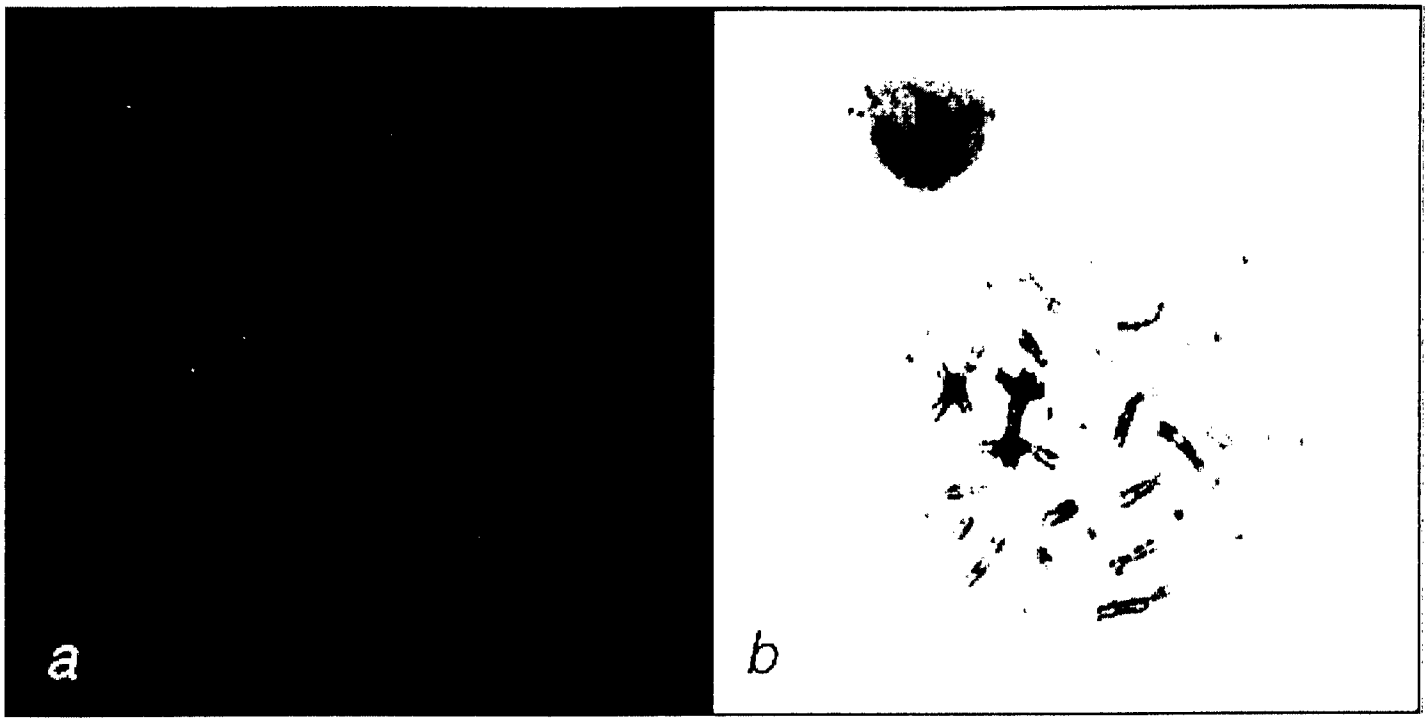


FIG. 3. Localization of a DOC-2-specific P1 clone, 35G12, using fluorescence *in situ hybridization* (FISH). (**a**) Illustration of a pseudocolored image of the localization of DOC-2 to band 5p13 of DAPI-counterstained normal male metaphase chromosomes. (**b**) Illustration of the same metaphase with pseudo-Giemsa-banded chromosomes.

chromosomal location and the genomic structure at the 5′-end of the gene. Sequence analysis on the DNA and protein levels revealed that DOC-2 is the human homologue of the murine mitogen-responsive phosphoprotein, p96, with 81% amino acid identity. The sequence comparisons also revealed that a 140 amino acid segment at the amino-terminal end of the gene, which in a separate study has been characterized as a potential PID, shares greater than 40% identity with genes isolated from distant species like *C. elegans* and *D. melanogaster*. Given the apparently ubiquitous expression pattern of DOC-2 in all cell types tested, together with the potential of p96 for phosphorylation at serine residues and the presence of the PID, we can hypothesize that DOC-2 is a tissue-independent component of cellular signal transduction.

ACKNOWLEDGMENTS

We gratefully acknowledge Dr. Suzi Mansour for providing mouse embryonic stem-cell RNA, Margaret Robertson and Elizabeth Lawrence from the DNA Sequencing Facility, Ed Meenen for synthesizing DNA oligonucleotides, and Ruth Foltz for editing the manuscript. Vysis generously provided the FISH analysis software. B.J.W. is supported by a fellowship from the American Foundation for Urologic Disease, and S.A.S. is supported by an EMBO fellowship. This work was supported by Grant DAM17-94-J-4129 from the Department of Defense to R.W.

REFERENCES

Albertsen, H. M., Smith, S. A., Mazoyer, S., Fujimoto, E., Stevens, J., Williams, B., Rodriguez, P., Cropp, C. S., Slijepcevic, P., Carlson, M., Robertson, M., Bradley, P., Lawrence, E., Sheng, Z. M., Hoopes, R., Sternberg, N., Brothman, A., Callahan, R., Ponder, B. A. J., and White, R. (1994). A physical map and candidate genes in the *BRCA1* region. *Nature Genet.* **7:** 472–479.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

Bork, P., and Margolis, B. (1995). A phosphotyrosine interaction domain. *Cell* **80:** 693–694 [Letter to the Editor].

Groden, J., Thliveris, A., Samowitz, W., Carlson, M., Gelbert, L., Albertsen, H., Joslyn, G., Stevens, J., Spirio, L., Robertson, M., Sargent, L., Krapcho, K., Wolff, E., Burt, R., Hughes, J. P., Warrington, J., McPherson, J., Wasmuth, J., Le Paslier, D., Abderrahim, H., Cohen, D., Leppert, M., and White, R. (1991). Identification and characterization of the familial adenomatous polyposis coli gene. *Cell* **66:** 589–600.

Kozak, M. (1991). An analysis of vertebrate mRNA sequences: Initiation of translational control. *J. Cell Biol.* **115:** 887–903.

Lee, W. H., Bookstein, R., Hong, F., Young, L. J., Shew, J. Y., and Lee, E. Y. (1987). Human retinoblastoma susceptibility gene: Cloning, identification, and sequence. *Science* **235:** 1394–1399.

Lichter, P., Chang, C-J. C., Call, K., Hermanson, G., Evans, G., Housman, D., and Ward, D. C. (1991). High resolution mapping of human chromosome 11 by in situ hybridization with cosmid clones. *Science* **247:** 64–69.

Liang, P., and Pardee, A. B. (1992). Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* **257:** 967–971.

Miki, Y., *et al.* (1994). A strong candidate gene for the breast and ovarian cancer susceptibility gene *BRCA1*. *Science* **266:** 66–71.

Mok, S. C., Wong, K-K., Chan, R. K. W., Lau, C. C., Tsao, S-W., Knapp, R., and Berkowitz, R. S. (1994). Molecular cloning of differentially expressed genes in human epithelial ovarian cancer. *Gynecol. Oncol.* **52:** 247–252.

Murakami, Y. S., Brothman, A. R., Leach, R. J., and White, R. L.

(1995). Suppression of malignant phenotype in a human prostate cancer cell line by fragments of normal chromosome 17q. *Cancer Res.* **55:** 3389–3394.

Shepherd, N. S., Pfrongner, B. D., Coulee, J. N., Ackerman, S. L., Vaidyanathan, G., Sauer, R. H., Balkenhol, T. C., and Sternberg, N. (1994). Preparation and screening of an arrayed human genomic library generated with the P1 cloning system. *Proc. Natl. Acad. Sci. USA* **91:** 2629–2633.

Smith, S. A., Holik, P. R., Stevens, J., Mazoyer, S., Melis, R., Williams, B., White, R., and Albertsen, H. (1996). Isolation of a gene encoding a second member of the disc-large family on chromosome 17q12–q21. *Genomics* **31:** 145–150.

Xu, X.-X., Yang, W., Jackowski, S., and Rock, C. O. (1995). Cloning of a novel phosphoprotein regulated by colony-stimulating factor 1 shares a domain with the Drosophila disabled gene product. *J. Biol. Chem.* **270:** 14184–14191.

Am. J. Hum. Genet. 56:484–499, 1995

A Strategy for Constructing High-Resolution Genetic Maps of the Human Genome: A Genetic Map of Chromosome 17p, Ordered with Meiotic Breakpoint–Mapping Panels

Steven C. Gerken,[1] Hans Albertsen,[1] Tami Elsner,[1] Linda Ballard,[1] Pilar Holik,[1] Elizabeth Lawrence,[1] Mary Moore,[1] Xuyun Zhao,[1] and Ray White[1,2]

[1]Department of Human Genetics and [2]Huntsman Cancer Institute, University of Utah, Salt Lake City

Summary

Genetic linkage analyses with genotypic data obtained from four CEPH reference families initially assigned 24 new PCR-based markers to chromosome 17 and located the markers at specific intervals of an existing genetic map of chromosome 17p. Each marker was additionally genotyped with an ordered set of obligate, phase-known recombinant chromosomes. The breakpoint-mapping panels for each family consisted of two parents, one sib with a nonrecombinant chromosome, and one or more sibs with obligate recombinant chromosomes. The relative order of markers was determined by sorting segregation patterns of new markers and ordered anchor markers and by minimizing double-recombination events. Consistency of segregation patterns with multiple flanking loci constituted support for order. A genetic map of chromosome 17p was completed with 39 markers in 23 clusters, with an average space of 3 cM between clusters. The collection of informative genotypes was highly efficient, requiring fivefold fewer genotypes than would be collected with all the CEPH families. Given the availability of large numbers of highly informative PCR-based markers, meiotic breakpoint mapping should facilitate construction of a human genomic map with 1-cM resolution.

Introduction

Recently published genetic linkage maps of human chromosomes are the result of an international effort to map and sequence the human genome (NIH/CEPH Collaborative Mapping Group 1992; Gyapay et al. 1994). These maps provide the information necessary for general linkage studies, including localizing genes with disease-causing alleles. Haplotype construction (Kelsell et al. 1994) and linkage disequilibrium studies (Bowcock et al. 1994; Jorde et al. 1994) in affected pedigrees can lead to high-resolution genetic maps over regions of a few centimorgans, but such studies are often limited by the number of informative meioses available. Advances in the positional cloning of genes have been rapid, since the advent of YAC libraries with cloned inserts of human genomic DNA (Burke et al. 1987). However, the lack of high-resolution genetic maps of the entire human genome means that new markers and new maps must be developed for each region surrounding a locus of interest, before the minimal physical region can be determined for molecular analysis. A genetic map of the human genome, at a resolution the same as or higher than that of current physical maps would facilitate integration of the physical and genetic maps and would accelerate cloning of the minimal region of the genome necessary to isolate any gene of interest. The physical map constructed by Cohen et al. (1993) provided the first large-scale effort to integrate a genetic map of the human genome with the physical map. Since some 10,000 genetic markers, ~4,500 of which are PCR based, have been reported to the Genome Data Base, it would appear that sufficient numbers of markers are available to develop a 1-cM genetic map spanning the 4,000–4,900 cM of the human genome (NIH/CEPH Collaborative Mapping Group 1992; Gyapay et al. 1994). The need for a high-resolution genetic map of the human genome will become increasingly urgent as genetic linkages are discovered for loci associated with the numerous human genetic disorders whose etiology is still unknown (McKusick 1992).

Multipoint linkage analysis with genotypic data collected from the CEPH reference families is a commonly used method for building genetic maps. To satisfy the statistical requirements of linkage analysis, genotyping of an extensive collection of individuals is necessary. An alternative approach to building genetic maps is meiotic breakpoint mapping, which utilizes known meiotic recombinant chromosomes to determine marker order on much smaller sample sizes (White et al. 1985; Fain et al. 1989; Weber et al. 1991; Maestri et al. 1992; Sprinkle et al. 1993; Attwood et al. 1994). The use of recombinant chromosomes as map-

Received July 26, 1994; accepted for publication November 17, 1994.

Address for reprints and correspondence: Steven C. Gerken, Ph.D., Department of Human Genetics, University of Utah, Salt Lake City, UT 84112

0002-9297/95/5602-0018$02.00

ping panels to construct high-resolution genetic maps has become routine for mapping murine loci and disease-causing genes in humans (European Backcross Collaborative Group 1994; Kelsell et al. 1994). In this approach, chromosomes with meiotic breakpoints known to reside between previously mapped loci are selected for genotyping. Segregation patterns of a test marker are compared with the patterns of anchor markers that flank the breakpoints, in a manner that minimizes the number of double-recombinant chromosomes. Segregation patterns of multiple test markers located in the region allow for the ordering of the test markers relative both to each other and to the known anchor markers. The rationale is that, although typing of all members of the CEPH panel would provide genotypic data for all of the meioses necessary to order new markers and for an estimate of the recombination fractions, using only the informative recombinant chromosomes localized to the region of the test locus constitutes a more efficient approach to ordering new markers.

Human chromosome 17 consists of 92 Mb of DNA, ~3% of the human genome. This chromosome is rich in loci involved in cancer; for example, it harbors the tumor-suppressor genes p53, NF1, and BRCA1, and loss of heterozygosity of other loci on chromosome 17p has been observed in breast carcinomas and uterine cancers (Isomura et al. 1994; Jones et al. 1994). Numerous other disease alleles have been mapped to this chromosome, including those responsible for the neurological diseases Charcot-Marie-Tooth type 1A (CMT1A) and hereditary neuropathy with liability to pressure palsies (Fain 1992; Chance and Pleasure 1993). More than 563 probes that detect polymorphic loci on chromosome 17, of which 227 are PCR based, are available. Almost 300 of these markers have been typed in the CEPH families (CEPH database, version 6), and, on the basis of these data, several genetic linkage maps of chromosome 17 have been developed (O'Connell et al. 1993; Gyapay et al. 1994).

The large number of available genetic markers and access to a genetic map composed of markers typed in 59 CEPH families made chromosome 17p an excellent choice for constructing a high-resolution genetic map of a whole chromosome arm, using recombinant-chromosome mapping panels. In this report, we describe a two-tiered approach to high-resolution mapping that reduces the number of genotypes necessary to order a new marker with respect to an existing map. Likelihood analysis was used to assign new markers to chromosome 17 and to determine the most likely intervals (defined by flanking index markers) for their locations on the index map. Using two new computer programs, RBUILD and CSORT (Elsner et al. 1995 [in this issue]), we constructed mapping panels of recombinant chromosomes and ordered new markers on the basis of analysis of the genotypic data collected from these mapping panels. Twenty-three distinct clusters, which comprised a total of 39 markers, were ordered for the 65-cM region of chromosome 17p, using mapping panels that represented a fivefold reduction in the number of genotypes collected.

Material and Methods

Genetic Markers

Twenty-four new PCR-based genetic markers, isolated for chromosome 17p, as part of a program to develop large numbers of PCR-based probes that detect human polymorphic loci, and three previously published PCR-based probes (2KZ14-B4, 2KZ22-B10, and UT172) that detected dinucleotide-repeat polymorphisms (table 1) (Utah Genome Center, personal communication) were used in this study. The new PCR-based probes were developed from the DNA sequences of clones containing microsatellite repeats detected by hybridization to end-labeled oligonucleotide probes d(CA)20, d(AAAG)20, d(AAAT)20, or d(AGC)20 (Melis et al. 1993). Twenty-six cloned DNA fragments from chromosome 17 that were used as molecular probes for detecting RFLPs or VNTRs were derived from previously published data (table 1).

Genotyping and Data Collection

Genotypes for the RFLP and VNTR probes were obtained from the data used to construct the genetic map of chromosome 17 (O'Connell et al. 1993). Autoradiographs of the original Southern blots were available for review for these markers, since all of the genotypes had been determined in our laboratory. Four of the CEPH reference families (1331, 1332, 1362, and 884) were genotyped with every PCR-based marker except UT18 and UT49, which were typed in families 1347, 1362, 1416, 1454, and 1463. Genotyping for the PCR-based markers was performed as described by Melis et al. (1993).

DNA for the mapping panels was aliquoted into master DNA trays, which were then dispensed in a Beckman Biomek 1000 pipetting robotic workstation to minimize sample errors. Genotypes were collected for each marker, with the mapping panel for which the most likely location of the test marker had been determined (fig. 1). For those loci whose most likely location showed no recombination with an index marker, genotypes were collected with the panel for the most distal interval. If the marker was not ordered within this interval, genotypes were collected with the panel for the next most likely interval.

Images of the autoradiographs were digitized with an image-capture system and were transferred to a computer-assisted genotyping system. All genotypes were determined independently by two individuals. Once genotypic data were verified, they were transferred to the Utah relational database for storage.

Meiotic Recombinant-Chromosome Panels

RBUILD, the computer program for selecting breakpoint-mapping panels, uses pairwise linkage analysis of

2KZ22B10[g] (D17S516)	L29355	*CCATTCCCATTGCTTCCATACTCA TGGAGTGGTCAGATCATATCACCT	GT_{16}	223	58/1.5	32	13	25	.48
UT65 (D17S748)	L36947	*CCAGGCCAATAATGCAACTCCTACTT GACCAGAGGAGACTCACAAAGGTTT	GT_{6}	119	65/3	19	13	32	.52
UT66 (D17S749)	L36947	*GGTAGCAAGCAGGTAGGACTTGG GAGCCAGTAAAAAAAAAGTCTGAG	$GT9GA_{15}$	112	60/1.5	11	7	5	.65
UT72 (D17S755)	L29356	*GAGTCCCAAAAGAGGATTTGGGC GTGACATGTTTTCCAAATCACATGT	GT_{9}	137	55/1.5	17	12	34	.76
UT137 (D17S596)	L29357	*CAAATTCAGGGGAGAAGGGACCA CCAGGACAGGAGTGTATCATCCA	GT_{6}	115	60/1.5	21	12	13	.49
UT146 (D17S606)	L29358	GGGCAACACAACAAGACCTTGTC *GGAATCCAGTTTCTCTCTGTAAGTA	$AAAT_{10}$	93	60/2	11	8	5	.63
UT158 (D17S619)	L29359	CCAGCACTGACCTTACAGGGTTT *TAAGAGATAGGCACAACCAACTGTT	GT_{11}	110	60/1	46	26	39	.51
UT159 (D17S620)	L29360	GGGAAGGTGTCTGAAACCCAAGG *CCCACCACTACCTATTGTTCTATAC	GT_{20}	143	60/1	22	11	17	.54
UT172[h] (D17S635)	L29361	GGTGAAAGAGCAAGACTCTGTCAC *CCCCTTGATTGTAAGCNACAGAAAC	$AAAT_{9}$	113	65/1.5	13	6	5	.50
UT184 (D17S647)	L29362	*ACTAGCAGCACGTTGTCTTTTGGG CACTGGATTTAAGTAGGGATGAGG	GT_{18}	94	60/1.5	19	13	23	.70
UT222 (D17S675)	L29363	CCAGGTGTAGTGACACATGCT *GTAGGCAGGTCAAGGTCTTAGGG	$AAAG_{25}$	193	60/1.5	23	13	8	.67
UT225 (D17S678)	L29370	CAGCTTGGCAACACAGCGAAA *TATTCTGCTCGGCACATAGTGCAA	AG_{15}	290	55/2	45	37	64	.80
UT263 (D17S689)	L29364	*TGCACCCAGCAATGGTGTCCTTT TCTAGCCTGGGTGACAGAGTGAA	$AAAG_{17}$	146	65/1.5	24	21	33	.86
UT265 (D17S691)	L29365	*GCCAGGGAGTGTACTACCTCCC AGCTGAGATTGTGCCACTTTG	$AAAG_{10}$	206	60/1	15	12	16	.80
UT269 (D17S695)	L29366	CTGGGCAACAAGAGCAAAATTC *TTTGTTGTTGTTCATTGACTTCAGTCT	$AAAG_{14}$	201	60/1.5	23	18	29	.91
UT403 (D17S1148)	L29372	AGGAAGGTTGAAGGCTGCA *TTCCCTGTTTCTGCCTAGG	$AAAG_{8}$	302	58/1.5	80	67	104	.89
UT405 (D17S900)	L16244	*TATAAGTGGTGGTTTCCCTG TGAGTGGAAAGTACAGAGAGT	CA_{15}	293	58/3	49	40	53	.85
UT751 (D17S906)	L16327	*AGCAAGATTCTGTCAAAAGAG TTCTAGCAGAGTGAAACTGTCT	$AAAG_{12}$	335	60/1	49	43	83	.91
UT1860 (D17S918)	L18709	TGTTGAGCTTTCCTGTAATC *TTCCTCACACAACCTATTGA	AGC_{8}	245	56/1	47	28	45	.61
UT1985 (D17S919)	L18714	*AGGCACAGAGTGAGACTTG GCTTAATTTTCACGAGGTTCAG	$GGAA_{15}$	148	58/1	80	66	99	.77
UT5265 (D17S1149)	L29374	AACAAGAGTGAACTCCATAGAGAG *CGCTGATCTGTCAGGCAGCCCT	$AAAG_{16}$	276	66/1.5	23	22	54	.92

[a] Locus number assigned by the Genome Database (DGB), Johns Hopkins University.
[b] Restriction enzyme used to reveal the RFLP or VNTR.
[c] Estimates of heterozygosity for the locus, determined in the CEPH families.
[d] Previously published by O'Connell et al. (1993).
[e] Haplotype of systems.
[f] Accession number assigned by the GenBank Sequence Database.
[g] Asterisk indicates the primer that was labeled with ^{32}P for visualization of the PCR product by autoradiography.
[h] Microsatellite repeat, determined by DNA-sequence analysis.
[i] Size of PCR product, as predicted by DNA-sequence analysis.
[j] Annealing temperature T (°C) and concentration of Mg^{++} (mM) used in the PCR reaction.
[k] Previously published by Shannon et al. (1994).

markers whose order is known, relative to the breakpoints, allows for orientation of the test markers, relative to the anchor markers.

Under a model of complete positive interference, the phase of each marker for each recombinant chromosome typed was determined by comparing the pattern of inheritance of the recombinant chromosome with that of the nonrecombinant chromosome. This pair of chromosomes is referred to as a "sibship pair." Sibship pairs were analyzed to determine whether the parental origin of each test marker was the same, different, or unknown, when compared with the nonrecombinant chromosome. The summary of comparisons for all markers for each recombinant chromosome (i.e., a "haplotype" of the parental contributions to each of these chromosomes) is referred to as the "chromosome composite" (Elsner et al. 1995).

The two markers for which the greatest number of sibship pairs are informative form the first interval of the breakpoint map. The recombinant sorting program, CSORT, tests the next-most-informative marker in each interval of the map, including the ends of the map, to determine whether a double-recombination event would be required for the marker to be located in this interval. The test marker is placed in the interval of the map where no double recombinant is required to explain the data. This process is iterated until all markers are tested. Markers that are excluded from all intervals are included in a secondary map. Next, the chromosome composites are shifted to order the breakpoints from right to left to provide an array of chromosome breakpoints whose order has been determined by the recombinant-sorting algorithm.

The foregoing technique detects possible genotypic errors in the data for a test marker when an anchor marker is ordered incorrectly or excluded from the breakpoint map. Autoradiographs for the test marker are then checked for errors. If placing a test marker has caused an anchor marker to be excluded or incorrectly ordered, it is removed from the breakpoint map and the analysis is restarted. Markers that cannot be ordered within the breakpoint map but whose location scores show high likelihoods for inclusion within an interval are reordered with only the anchor markers for the interval. Segregation patterns that exclude the marker from certain locations and that minimize the number of double or multiple recombinants determine the location of the test marker.

Sequence Analysis of Sequence-Tagged Sites (STS)

DNA-sequence databases representing all STS-containing microsatellite repeats from the Utah, Généthon, CHLC, and Marshfield collections were constructed using the GCG (1991) program DATASET. DNA sequences for the Généthon, Marshfield, and CHLC markers were retrieved from the GenBank sequence database (version 78) or the CHLC server. DNA sequence analyses for STS with significant identity to the Utah microsatellites were performed with the GCG program QUICKSEARCH, and results were displayed with the program QUICKSHOW.

Table 2

Pairwise Analysis of Linkage Between 24 UTAH Markers and Généthon Markers: Lod Score Z_{max} at Recombination Fraction Theta

UTAH Marker	CEPH Marker	Theta	Z_{max}
UT18	D17S849	.0010	4.8
UT20	D17S796	.056	12.8
UT25	D17S783	.001	13.5
UT39	D17S796	.001	15.6
UT49	D17S796	.001	5.4
UT65	D17S786	.043	9.4
UT66	D17S798	.001	6.0
UT72	D17S786	.024	10.6
UT137	D17S804	.001	6.3
UT146	D17S798	.001	12.3
UT158	D17S786	.047	8.2
UT159	D17S783	.001	6.6
UT184	D17S796	.037	9.5
UT222	D17S786	.050	8.3
UT225	D17S796	.012	20.8
UT263	D17S783	.001	12.6
UT263	D17S805	.001	12.6
UT269	D17S849	.001	12.6
UT403	D17S786	.040	11.7
UT405	D17S799	.026	9.4
UT751	D17S796	.001	19.8
UT1860	D17S783	.001	12.6
UT1985	D17S796	.001	7.2
UT5265	D17S796	.001	13.8

Results

Genetic Markers

Twenty-four new PCR-based probes were assigned to chromosome 17 (table 1) on the basis of results obtained by two-point linkage analysis (table 2). The microsatellite repeats detected by these probes included 10 d(GT), 8 d(AAAG), 3 d(AAAT), and other motifs (table 1). DNA sequences of the primer pairs, the primers to end-label for visualization by autoradiography, and the optimized conditions for PCR are reported in table 1. Accession numbers are provided to facilitate retrieval of the DNA sequence of each STS from the GenBank sequence database.

Estimates of heterozygosity for the polymorphic loci ranged from .49 to .91, with an average heterozygosity of .68. Heterozygosities for the dinucleotide repeat–containing loci averaged .65; and those for tetranucleotide repeat–containing loci d(AAAT) and d(AAAG) averaged .59 and .84, respectively. The autoradiographic images obtained with PCR probes for the loci containing tri- and

Table 3

Meiotic Recombinant-Chromosome Mapping Panels

CEPH Family[a]	Chromosomes Selected[b]
Obligate recombinant and nonrecombinant chromosomes for the interval 144D6–LB17.8:	
13281 Recombinant	*3 paternal
13281 Nonrecombinant	*4 paternal
13291 Recombinant	*9 paternal
13291 Nonrecombinant	*8 paternal
13293 Recombinant	3 paternal
13293 Recombinant	*7 paternal
13293 Recombinant	*9 paternal
13293 Nonrecombinant	*8 paternal
13294 Recombinant	6 paternal
13294 Recombinant	6 maternal
13294 Nonrecombinant	3 paternal
13294 Nonrecombinant	3 maternal
1331 Recombinant	*6 paternal
1331 Recombinant	*7 paternal
1331 Recombinant	*16 paternal
1331 Nonrecombinant	*3 paternal
1332 Recombinant	*8 paternal
1332 Nonrecombinant	*6 paternal
1333 Recombinant	*8 maternal
1333 Recombinant	10 paternal
1333 Recombinant	10 maternal
1333 Nonrecombinant	3 paternal
1333 Nonrecombinant	*3 maternal
1340 Recombinant	*3 paternal
1340 Recombinant	4 paternal
1340 Recombinant	*7 paternal
1340 Recombinant	*8 paternal
1340 Nonrecombinant	*13 paternal
1341 Recombinant	*3 paternal
1341 Recombinant	*8 paternal
1341 Nonrecombinant	*4 paternal
1346 Recombinant	*7 paternal
1346 Nonrecombinant	*8 paternal
1347 Recombinant	*3 maternal
1347 Recombinant	*10 maternal
1347 Nonrecombinant	*4 maternal
1349 Recombinant	*3 paternal
1349 Recombinant	4 paternal
1349 Recombinant	6 maternal
1349 Nonrecombinant	*8 paternal
1349 Nonrecombinant	5 maternal
1350 Recombinant	*7 paternal
1350 Nonrecombinant	*5 paternal
1353 Recombinant	15 paternal
1353 Nonrecombinant	5 paternal
1354 Recombinant	*3 maternal
1354 Recombinant	*4 paternal
1354 Nonrecombinant	*3 paternal
1354 Nonrecombinant	*4 maternal
1355 Recombinant	10 paternal
1355 Recombinant	13 paternal
1355 Nonrecombinant	3 paternal
1357 Nonrecombinant	*14 paternal
1357 Nonrecombinant	4 maternal
1362 Recombinant	*4 paternal
1362 Recombinant	*5 paternal
1362 Nonrecombinant	*3 paternal
1413 Recombinant	*9 paternal
1413 Recombinant	*14 paternal
1413 Nonrecombinant	*12 paternal
1416 Recombinant	4 paternal
1416 Recombinant	6 paternal
1416 Recombinant	10 paternal
1416 Nonrecombinant	8 paternal
1418 Nonrecombinant	*9 paternal
1418 Nonrecombinant	*8 paternal
1421 Recombinant	7 maternal
1421 Recombinant	*12 paternal
1421 Nonrecombinant	*11 paternal
1421 Nonrecombinant	3 maternal
1423 Recombinant	*3 paternal
1423 Recombinant	*9 paternal
1423 Nonrecombinant	*8 paternal
66 Recombinant	3 maternal
66 Recombinant	4 maternal
66 Recombinant	6 paternal
66 Recombinant	8 paternal
66 Nonrecombinant	3 paternal
66 Nonrecombinant	5 maternal
104 Recombinant	4 paternal
104 Recombinant	8 paternal
104 Recombinant	10 paternal
104 Nonrecombinant	5 paternal
1444 Recombinant	*9 paternal
1444 Recombinant	*10 paternal
1444 Nonrecombinant	*8 paternal
1447 Recombinant	*6 maternal
1447 Recombinant	7 paternal
1447 Recombinant	*8 maternal
1447 Nonrecombinant	3 paternal
1447 Nonrecombinant	*7 maternal
1454 Recombinant	*4 paternal
1454 Nonrecombinant	*7 paternal
1458 Recombinant	5 paternal
1458 Nonrecombinant	3 paternal
1459 Recombinant	*3 paternal
1459 Recombinant	*5 paternal
1459 Recombinant	7 maternal
1459 Nonrecombinant	*4 paternal
1459 Nonrecombinant	4 maternal
1463 Recombinant	*4 paternal
1463 Recombinant	*7 paternal
1463 Recombinant	*10 paternal
1463 Recombinant	*12 paternal

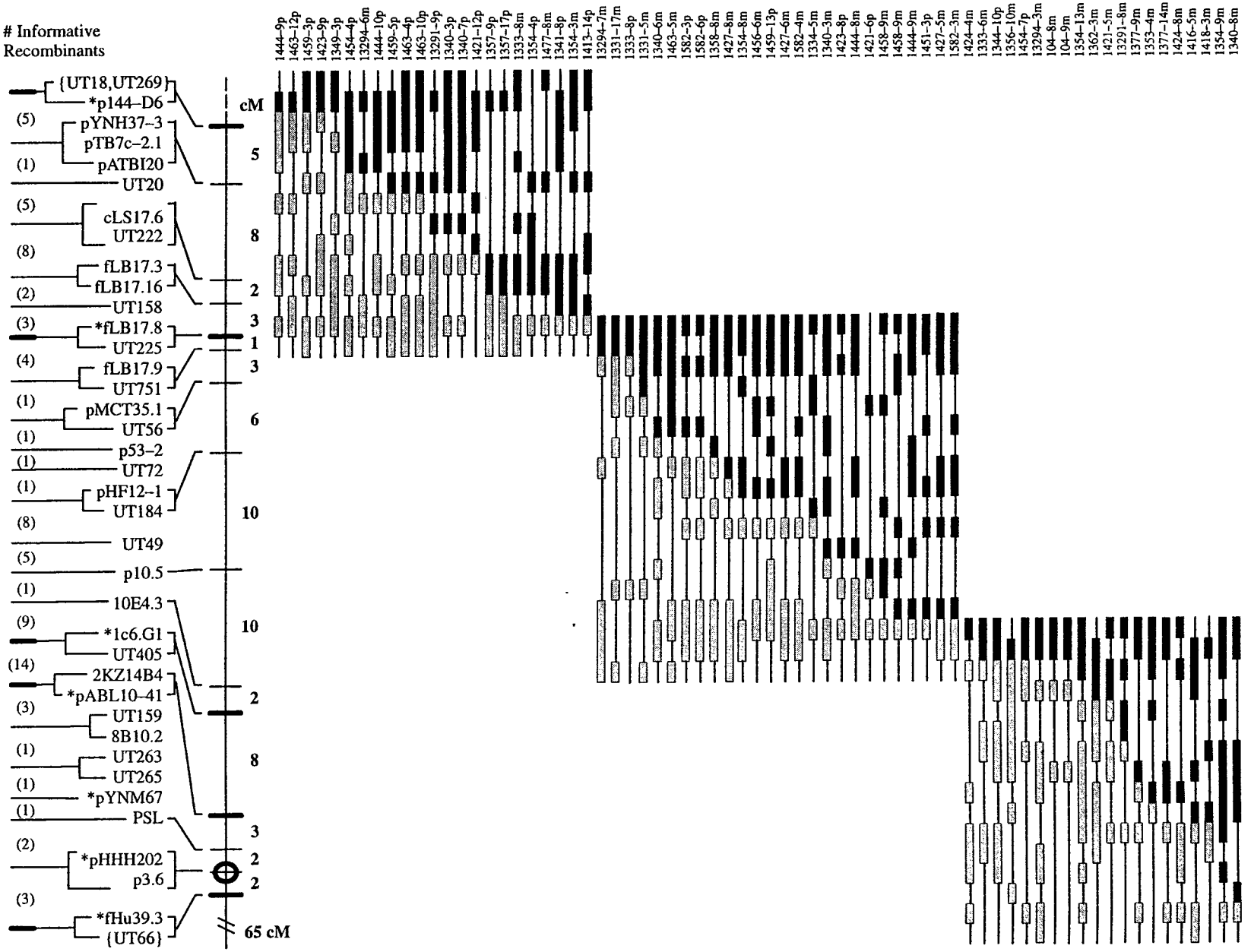
(*continued*)

Table 3 (continued)

CEPH Family[a]	Chromosomes Selected[b]
Obligate recombinant and nonrecombinant chromosomes for the interval LB17.8–1C6G1: (*continued*)	
1358 Nonrecombinant	*9 maternal
1362 Recombinant	*12 paternal
1362 Nonrecombinant	*5 paternal
1377 Recombinant	3 maternal
1377 Nonrecombinant	4 maternal
1408 Recombinant	4 maternal
1408 Recombinant	8 maternal
1408 Nonrecombinant	6 maternal
1413 Recombinant	*7 paternal
1413 Recombinant	*10 paternal
Obligate recombinant and nonrecombinant chromosomes for the interval 1C6G1–fHu39.3:	
13281 Recombinant	*4 maternal
13281 Nonrecombinant	*5 maternal
13291 Recombinant	*6 maternal
13291 Recombinant	*8 maternal
13291 Nonrecombinant	*3 maternal
13294 Recombinant	*3 maternal
13294 Nonrecombinant	*4 maternal
1333 Recombinant	*6 maternal
1333 Recombinant	*7 maternal
1333 Nonrecombinant	*3 maternal
1340 Recombinant	*8 maternal
1340 Nonrecombinant	*3 maternal
1344 Recombinant	*9 maternal
1344 Recombinant	10 paternal
1344 Recombinant	11 maternal
1344 Nonrecombinant	3 paternal
1344 Nonrecombinant	*3 maternal
1347 Recombinant	*7 maternal
1347 Recombinant	*11 maternal
1347 Nonrecombinant	*3 maternal
1353 Recombinant	*4 maternal
1353 Nonrecombinant	*5 maternal
1354 Recombinant	*9 maternal
1354 Recombinant	*13 maternal
1354 Nonrecombinant	*3 maternal
1356 Recombinant	*10 maternal
1356 Recombinant	*15 maternal
1356 Nonrecombinant	*6 maternal
1358 Recombinant	*4 maternal
1358 Recombinant	*5 maternal
1358 Recombinant	*9 paternal
1358 Recombinant	*9 maternal
1358 Recombinant	*10 maternal
1358 Nonrecombinant	*3 paternal
1400 Recombinant	5 maternal
1400 Nonrecombinant	3 maternal
1427 Recombinant	*5 maternal
1427 Recombinant	*6 maternal
1427 Recombinant	*8 maternal
1427 Nonrecombinant	*7 maternal
1582 Recombinant	*3 paternal
1582 Recombinant	*3 maternal
1582 Recombinant	*4 maternal
1582 Recombinant	*6 paternal
1582 Recombinant	*9 paternal
1582 Nonrecombinant	*4 paternal
1582 Nonrecombinant	*6 maternal
1358 Nonrecombinant	*3 maternal
1362 Recombinant	*3 maternal
1362 Nonrecombinant	*5 maternal
1377 Recombinant	8 paternal
1377 Recombinant	*9 maternal
1377 Recombinant	*14 maternal
1377 Nonrecombinant	*3 maternal
1416 Recombinant	*4 maternal
1416 Recombinant	*5 maternal
1416 Nonrecombinant	*3 maternal
1418 Recombinant	*3 maternal
1418 Nonrecombinant	*6 maternal
1421 Recombinant	*5 maternal
1421 Nonrecombinant	*3 maternal
1423 Recombinant	*3 maternal
1423 Nonrecombinant	*4 maternal
1424 Recombinant	*4 maternal
1424 Recombinant	*8 maternal
1424 Nonrecombinant	*5 maternal
104 Recombinant	*8 maternal
104 Recombinant	*9 maternal
104 Nonrecombinant	*5 maternal
1447 Recombinant	*3 maternal
1447 Nonrecombinant	*4 maternal
1454 Recombinant	7 paternal
1454 Nonrecombinant	3 paternal
1456 Recombinant	*10 maternal
1456 Nonrecombinant	*4 maternal
1458 Recombinant	*8 maternal
1458 Nonrecombinant	*4 maternal
1427 Recombinant	*4 maternal
1427 Nonrecombinant	*9 maternal
1477 Recombinant	*7 paternal
1477 Nonrecombinant	*3 paternal

[a] "Recombinant" indicates that the specified sibling contains an obligate recombinant chromosome inherited from the specified parent; and "Nonrecombinant" indicates that the specified sibling contains a nonrecombinant chromosome inherited from the specified parent.

[b] In specified sibling. An asterisk (*) indicates an individual typed with PCR-based test markers.


Informative Recombinants
{UT18,UT269}
*p144–D6
(5) pYNH37–3
pTB7c–2.1
(1) pATBI20
UT20
(5) cLS17.6
UT222
(8) fLB17.3
fLB17.16
(2) UT158
(3) *fLB17.8
UT225
(4) fLB17.9
UT751
(1) pMCT35.1
UT56
(1) p53–2
(1) UT72
(1) pHF12–1
UT184
(8) UT49
(5) p10.5
(1) 10E4.3
(9) *1c6.G1
UT405
(14) 2KZ14B4
*pABL10–41
(3) UT159
8B10.2
(1) UT263
UT265
(1) *pYNM67
(1) PSL
(2) *pHHH202
p3.6
(3) *fHu39.3
{UT66}
cM
5
8
2
3
1
3
6
10
10
2
8
3
2
2
65 cM
1444–9p
1463–12p
1459–3p
1423–9p
1349–3p
1454–4p
13294–6m
1444–10p
1459–5p
1463–4p
1463–10p
13291–9p
1340–3p
1340–7p
1421–12p
1357–9p
1357–17p
1333–8m
1354–4p
1477–8m
1341–8p
1354–3m
1413–14p
13294–7m
1331–17m
1333–8p
1331–5m
1340–6m
1463–5m
1582–3p
1582–6p
1358–8m
1427–8m
1354–8m
1456–6m
1459–13p
1427–6m
1582–4m
1334–5m
1340–3m
1423–8p
1444–8m
1421–6p
1458–9m
1458–9m
1444–9m
1451–3p
1427–5m
1582–3m
1424–4m
1333–6m
1344–10p
1356–10m
1454–7p
13294–3m
104–8m
104–9m
1354–13m
1362–3m
1421–5m
13291–8m
1377–9m
1353–4m
1377–14m
1424–8m
1416–5m
1418–3m
1354–9m
1340–8m

one obligate recombinant; the multipoint map did not report the relative order of these two loci (Fain 1992). The orders of all markers flanking the CMT1A locus that were common to both maps were also identical. For loci flanking the NF1 locus, O'Connell et al. (1989) reported 1% recombination between p3.6 and pHHH202, but we observed no recombination between these loci. This discrepancy reflected the fact that the recombinant in kindred 1333, individual 9, which was observed in the study by O'Connell et al. (1989), was not detected by the anchor markers used to identify obligate recombinants in this study. UT172 was reported to be located 1.5 Mb proximal to NF1 (Shannon et al. 1994). Our recombinant panel also located UT172 proximal to NF1 (table 5).

An assumption of complete positive interference in the small genetic intervals analyzed here is supported by recent evidence of interference in human meioses (Kwiatkowski et al. 1993; Weber et al. 1993). Interference minimizes multiple-recombination events and makes the breakpoint-panel approach to ordering genetic loci more feasible (Copeland and Jenkins 1991). Observed double-recombinant events in our analysis of breakpoints on chromosome 17p were most likely due to genotypic errors or to new undetected mutations that converted alleles to those already occurring in the family. Errors in the collection and interpretation of genotypic data occur at a frequency of ~1% in the CEPH database (Buetow 1991), while estimates of meiotic gene conversions for STRP without recombination have been estimated at 1:3,300 meioses (Weber et al. 1993). While the observed "double recombinants" are likely to be genotyping errors, they may also provide evidence of events such as gene conversion. Although we were careful to minimize the occurrence of errors in genotypic data, we would need to retype the family members in question, using DNA from primary sources such as blood, to better understand the nature of the "double-recombinant" chromosomes.

This approach to construction of high-resolution genetic maps decreased the collection of genotypic data from 899 individuals for the extended reference families to ~180 individuals for the present study. In the past, meiotic recombinant panels have been used only for determining the order of a few genetic loci, especially those with disease-causing alleles. The approach and tools presented here allow for a large number of loci and meiotic recombinant chromosomes to be ordered in seconds of computer time and is similar to that used to construct high-resolution linkage maps of the murine genome (Guenet and Brown 1993). The incorporation of "anchor markers," for which order has already been well established, provides a level of confidence that significant errors have not been introduced into the map. Simultaneous evaluation of large numbers of genetic markers typed in the meiotic chromosome–mapping panels is a logical use of the computational capacity of the computer.

The achievement of a genetic map of chromosome 17p, with 39 markers organized into 23 clusters with an average spacing of 3 cM between clusters, provides a level of resolution that can be useful in either the construction of or comparison with the physical map. When the physical and genetic maps are integrated at such a high resolution, a genetic map can be scaled in megabases, and it will be possible to compare recombination fractions with physical distances within the genome. For regions with unusual recombination features (i.e., high recombination fractions over small physical distances), the boundaries of the recombination events will also have been mapped in the CEPH families.

The clusters of loci where the markers are not separated by recombination events could be due to the fact that the appropriate recombinant chromosomes necessary to order these loci were not genotyped or that the recombination events are not randomly distributed throughout this arm of chromosome 17p. Differences in sex-specific recombination at various regions of human and murine chromosomes have been reported (O'Connell et al. 1987; Shiroishi et al. 1991). Determination of the distribution of the chromosome 17p recombination sites, with probes for large numbers of polymorphic loci, will be necessary to test whether the distribution of recombination events in the CEPH chromosomes are random or clustered.

The isolation by different laboratories, of the same microsatellite repeat containing loci reflects the fact that the number of polymorphic loci in the human genome is finite. Although the number of dinucleotide repeat–containing loci in the human genome is an estimated several hundred thousand, bias introduced by the use of biological cloning vectors, the size selection of inserts, and the length of microsatellite repeats influence the population of loci that will be isolated. DNA sequence analysis should be performed to determine whether newly isolated loci detected by STS are identical to existing loci. This method offers a new tool for the geneticist to determine whether two genetic loci are identical and represents another approach to determining whether errors exist in genotypic data.

Our observation that UT72 (D17S755) is identical in sequence to the Généthon marker D17S786 serves to anchor the two genetic maps of chromosome 17p at this locus. UT72, which is <28 cM from the most distal marker in this study, 144-D6, is reported to be 19 cM from the most distal marker on the Généthon map. Sequence-known loci should expedite the integration of the physical and genetic maps, a notion proposed when the STS was adopted as the "common language" of the human genome program (Olsen et al. 1989).

If the genetic loci detected by the newly mapped probes were uniformly distributed and fully informative,

O'Connell P, Plaetke R, Matsunami N, Odelberg S, Jorde L, Chance P, Leppert M, et al (1993) An extended genetic linkage map and an "index" map for human chromosome 17. Genomics 15:38–47

Olsen M, Hood L, Cantor C, Botstein D (1989) A common language for the physical map of the human genome. Science 245:1434–1435

Shannon KM, O'Connell P, Martin GA, Paderanga D, Olson K, Dinndorf P, McCormick F (1994) Loss of the normal NF1 allele from the bone-marrow of children with type 1 neurofibromatosis and malignant myeloid disorders. N Engl J Med 330:597–601

Shiroishi T, Sagai T, Hanzawa N, Gotoh H, Moriwaki K (1991) Genetic control of sex-dependent meiotic recombination in the major histocompatibility complex of the mouse. EMBO J 10:681–686

Sprinkle TJ, Kouri RE, Fain PD, Stoming TA, Whitney JB III (1993) Chromosomal mapping of the human CNP gene using a meiotic crossover DNA panel, PCR and allele-specific probes. Genomics 16:542–545

Vanagaite L, Savitsky K, Rotman G, Ziv Y, Gerken SC, White R, Weissenbach J, et al (1994) Physical localization of microsatellite markers at the ataxia-telangiectasia locus at 11q22-23. Genomics 22:231–233

Weber JL, Polymeropolous MH, May PE, Kwitek AE, Xiao H, McPherson JD, Wasmuth JJ (1991) Mapping of human chromosome 5 microsatellite DNA polymorphisms. Genomics 11:695–700

Weber JL, Wang Z, Hansen K, Stephenson M, Kappel C, Salzman S, Wilkie PJ, et al (1993) Evidence for human meiotic recombination interference obtained through construction of a short tandem repeat–polymorphism linkage map of chromosome 19. Am J Hum Genet 53:1079–1095

White R, Lalouel J-M (1987) Interference and mapping function. In: Harris G, Hirschhorn K (eds) Advances in human genetics, vol 16. Plenum, New York, pp 121–228

White R, Leppert M, Bishop DT, Barker D, Berkowitz J, Brown C, Callahan P, et al (1985) Construction of linkage maps with DNA markers for human chromosomes. Nature 313:101–105

Inhibition of DNA methyltransferase stimulates the expression of signal transducer and activator of transcription 1, 2, and 3 genes in colon tumor cells

Adam R. Karpf, Peter W. Peterson, Joseph T. Rawlins, Brian K. Dalley, Qian Yang, Hans Albertsen, and David A. Jones*

The Huntsman Cancer Institute, University of Utah, Salt Lake City, UT 84112

Communicated by Raymond L. White, University of Utah, Salt Lake City, UT, September 23, 1999 (received for review May 5, 1999)

Inhibitors of DNA methyltransferase, typified by 5-aza-2′-deoxycytidine (5-Aza-CdR), induce the expression of genes transcriptionally down-regulated by *de novo* methylation in tumor cells. We utilized gene expression microarrays to examine the effects of 5-Aza-CdR treatment in HT29 colon adenocarcinoma cells. This analysis revealed the induction of a set of genes that implicated IFN signaling in the HT29 cellular response to 5-Aza-CdR. Subsequent investigations revealed that the induction of this gene set correlates with the induction of signal transducer and activator of transcription (STAT) 1, 2, and 3 genes and their activation by endogenous IFN-α. These observations implicate the induction of the IFN-response pathway as a major cellular response to 5-Aza-CdR and suggests that the expression of STATs 1, 2, and 3 can be regulated by DNA methylation. Consistent with STAT's limiting cell responsiveness to IFN, we found that 5-Aza-CdR treatment sensitized HT29 cells to growth inhibition by exogenous IFN-α2a, indicating that 5-Aza-CdR should be investigated as a potentiator of IFN responsiveness in certain IFN-resistant tumors.

DNA cytosine methyltransferase I (DNA MeTase) recognizes hemimethylated CpG dinucleotides in mammalian DNA and catalyzes the transfer of methyl groups to cytosine residues in newly synthesized DNA (1). The methylation of cytosines within CpG islands located in core promoter regions can negatively regulate the transcription of the adjacent genes. The basis for this negative regulation may involve recruitment of histone deacetylases to methylated CpG islands (1). Holliday first suggested a relationship between abnormal DNA methylation and cancer (2). Subsequently, a number of methylation-silenced tumor suppressor genes, including p16Ink4a, retinoblastoma, estrogen receptor, hMLH1, and E-cadherin, have been identified in cancer cells *in vitro* and *in vivo* (3–8). It is becoming clear that epigenetic processes constitute a significant factor in the formation of cancer (9). In this regard, DNA methylation abnormalities have been implicated in colon cancers in both mouse and human tumor model systems (6, 10–12).

5-Aza-2′-deoxycytidine (5-Aza-CdR) inhibits DNA methylation and often is used *in vitro* to induce the reexpression of genes putatively silenced by promoter methylation (8). 5-Aza-CdR is substituted for cytosine during replication and is recognized by DNA MeTase (13). Attempted transfer of methyl groups to 5-Aza-CdR, however, covalently traps the enzyme to newly synthesized DNA (14, 15). This sequestration ultimately depletes cellular stores of DNA MeTase and results in widespread genomic hypomethylation. Clinical trials have demonstrated promise in the use of 5-Aza-CdR (decitabine) for treating leukemia, and current trials are evaluating 5-Aza-CdR in the treatment of lung and prostate cancers (16–19). It is plausible that the antitumor activity of 5-Aza-CdR results from the induction of methylation-regulated tumor-suppressive pathways.

The identification of methylation-silenced genes is offering new insights into tumor development and may reveal the potential for inhibiting DNA methylation as a cancer treatment (20). In this regard, a number of strategies have been used to uncover methylation-regulated genes, including candidate gene analysis, representational difference analysis, restriction landmark genome scanning, methylation-sensitive, arbitrarily primed PCR, and methylated DNA-binding protein affinity chromatography (21–24). Another strategy, gene expression microarrays, is particularly suited for identifying candidate, methylation-silenced genes and for assessing the downstream, cellular consequences of reactivating these genes. Microarray technology permits the systematic examination of thousands of gene expression changes simultaneously and has been used to follow the transcriptional changes that accompany disease development and cellular responses to environmental stimuli (25–29).

In view of the clinical interest in 5-Aza-CdR and our incomplete understanding of the cellular consequences of inhibiting DNA MeTase, we have utilized gene expression microarrays to probe the effects of treating colon tumor cells with 5-Aza-CdR. Here, we show that 5-Aza-CdR inhibits the growth of HT29 colon carcinoma cells and that this growth inhibition parallels the transcriptional induction of IFN-responsive genes. Subsequent analysis revealed induction of signal transducers and activators of transcription 1, 2, and 3 (STATs 1, 2, and 3), elements central to IFN signaling. Given the established growth-inhibitory properties of IFNs, our data offer a new model for understanding the cellular consequences of inhibiting DNA MeTase.

Materials and Methods

Cell Culture and Drug Treatments. HT29 adenocarcinoma cells (American Type Culture Collection) were cultured at 37°C in 5% CO_2 by using McCoy's medium supplemented with 10% FBS (GIBCO). For treatments with 5-Aza-CdR, cells were exposed to 500 nM 5-Aza-CdR (Sigma) 24 hr after passage in complete culture medium. Control cultures were treated in parallel with vehicle (PBS). Twenty-four hours after drug addition, culture medium was replaced with drug-free medium. Control and 5-Aza-CdR-treated cells were subcultured at equal densities at 1 and 5 days after the initial treatment, and proliferation was measured at the subsequent time point by using a Coulter counter.

In other experiments, HT29 cells (control or pretreated with 500 nM 5-Aza-CdR) were exposed to human recombinant IFN-α2a (a gift from Roche) at 1×10^5 units/ml or human recombinant IFN-γ (GIBCO) at 5×10^2 units/ml. RNA was harvested for microarray expression analysis at 10, 24, and 96 hr. IFN concentrations were established by measuring growth inhibition in HT29 cells after treatment and approximating the IC_{50} for each IFN type (data not shown).

Abbreviations: 5-Aza-CdR, 5-aza-2′-deoxycytidine; DNA MeTase, DNA cytosine methyltransferase; STAT, signal transducer and activator of transcription; EST, expressed sequence tag.

*To whom reprint requests should be addressed at: University of Utah, 2000 Circle of Hope, Salt Lake City, UT 84112. E-mail: david.jones@hci.utah.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Construction of Microarrays. The cDNA clones on the microarray were obtained from Research Genetics (Huntsville, AL) and Genome Systems (St. Louis). Transformants were grown overnight at 37°C in 96-well microtiter dishes containing 0.2 ml/well Terrific Broth supplemented with ampicillin. Cultures were transferred to a Millipore multiscreen, 96-well glass-fiber filtration plate (MAFB NOB), and growth medium was voided. Twenty-five microliters of 25 mM Tris·HCl, pH 8/10 mM EDTA/50 μl of 0.2 M NaOH/1% SDS/160 μl of 0.7 M potassium acetate, pH 4.8/5.3 M guanidine hydrochloride was added to each well of the glass filtration plate. Cell lysates were drawn through the glass filters under vacuum, and filter-bound DNA was washed four times with 200 μl of 80% ethanol. Plasmid DNAs were eluted by centrifugation after the addition of 65 μl of distilled H_2O. Samples were collected in a 96-well microtiter dish during centrifugation.

PCR amplifications (30 cycles, 52°C annealing) were performed in 100-μl reaction volumes in a 96-well format by using 2 μl of purified plasmid as template and vector-specific primers (typically T7 and T3). PCR products were combined with 200 μl of binding solution (150 mM potassium acetate, pH 4.8/5.3 M guanidine hydrochloride) in a Millipore multiscreen glass-fiber filtration plate. Vacuum was applied to void the binding solution, and bound PCR product was washed four times with 200 μl of 80% ethanol. Products were eluted in 65 μl of distilled H_2O. PCR product size ranged from 300 bp to 2.0 kb, with 1.0 kb as a typical length. DNA was prepared for spotting by diluting the purified PCR products in DMSO at a final concentration of 20–45 ng/μl.

Microarray slides were produced by using a Generation III Microarray Spotter (Molecular Dynamics). Each microarray contained 4,608 minimally redundant cDNAs spotted in duplicate on 3-aminoproply-trimethoxy silane-coated (Sigma) slides and UV crosslinked in a Stratalinker (Stratagene).

Generation of Microarray Probes, Microarray Hybridizations, and Scanning. Total RNA was isolated by using Trizol reagent (GIBCO) and poly(A) RNA was selected by using an Oligotex Kit (Qiagen). First-strand cDNA probes were generated by incorporation of Cy3-dCTP or Cy5-dCTP (Amersham Pharmacia) during reverse transcription of purified mRNA (1 μg) with SuperScript II (GIBCO). After synthesis, RNA/cDNA hybrids were denatured and the mRNA was hydrolyzed with NaOH. Single-stranded cDNA probes were transferred to a Millipore glass-fiber filtration plate containing two volumes of 150 mM potassium acetate, pH 4.8, and 5.3 M guanidine hydrochloride. The mixture was voided by vacuum, and bound cDNA was washed four times with 80% ethanol. Probes were eluted by the addition of 50 μl of distilled H_2O, recovered by vacuum concentration, and reconstituted in 30 μl of 5× SSC/0.1% SDS/0.1 μg/ml salmon sperm DNA/50% formamide. After denaturation at 94°C, the hybridization mixture was deposited onto an arrayed slide under a coverslip.

Hybridizations were performed overnight at 42°C in a humidified chamber. After hybridization, slides were washed for 10 min in 1× SSC/0.2% SDS and then for 20 min in 0.1× SSC/0.2% SDS. Slides were rinsed briefly in distilled water and dried with compressed air, and the fluorescent hybridization signatures were captured by using the "*Avalanche*" dual-laser confocal scanner (Molecular Dynamics). Fluorescent intensities were quantified by using ARRAYVISION 4.0 (Imaging Research, St. Catherine's, ON, Canada).

Northern Blotting and Reverse Transcription–PCR. Five micrograms of total RNA was fractionated through formaldehyde-containing agarose gels and transferred onto nylon membranes (Amersham Pharmacia). Hybridizations with ^{32}P-labeled probes were carried out by using Rapid-hyb buffer (Amersham Pharmacia). Reverse transcription–PCR of type I (α, β) and II (γ) IFN genes was carried out on cDNAs prepared from vehicle-treated and 500 nM 5-Aza-CdR-treated HT29 cells 9 days after treatment. The primers used for amplification of IFN-α are within the coding region and are capable of amplifying each member of the IFN-α gene cluster.

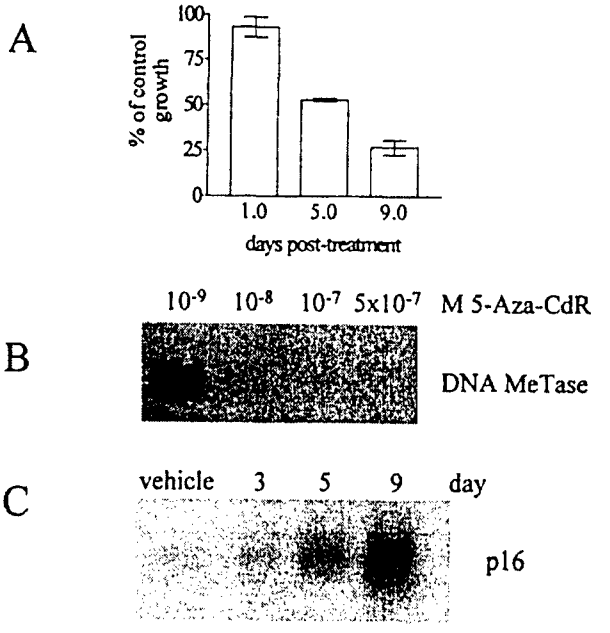


Fig. 1. 5-Aza-CdR inhibits HT29 cell proliferation and sequesters DNA MeTase I. (*A*) HT29 cells were treated with vehicle or 500 nM 5-Aza-CdR for 24 hr. After this treatment, the drug was removed and cell proliferation was measured by directly counting cells at the indicated time points (see *Materials and Methods*). Data are presented as mean count ± 1 SD, ($n = 3$). (*B*) HT29 cells were treated with the indicated concentrations of 5-Aza-CdR for 24 hr, and the presence of DNA MeTase I (200 kDa) was assessed in nuclear protein extracts by immunoblotting. Sequestration of DNA MeTase I by 500 nM 5-Aza-CdR continued for 4 days after treatment (data not shown) (*C*) The expression of p16 in HT29 cells at time points after treatment with 500 nM 5-Aza-CdR was measured by Northern blot analysis.

Cell Fractionations and Western Blotting. Nuclear and cytoplasmic fractions were prepared as described previously (30). Protein extracts (50 μg) were fractionated through 10% SDS/PAGE gels (Novex) and blotted onto poly(vinylidene difluoride) membranes (Amersham Pharmacia). Antibody to DNA methyltransferase I was a kind gift from Moshe Szyf (McGill University, Montreal, Canada). STATs 1, 2, and 3 antibodies were purchased from Transduction Laboratories (Lexington, KY). Final protein detection employed a horseradish peroxidase-conjugated goat anti-mouse secondary antibody (GIBCO) and chemiluminescence (NEN Renaissance).

Results

5-Aza-CdR Treatment Inhibits the Proliferation of HT29 Cells. HT29 colon adenocarcinoma cells are p53- and APC-deficient and mismatch repair-proficient. Treatment of these cells with 500 nM 5-Aza-CdR for 24 hr caused a time-dependent, 3-fold inhibition of proliferation (Fig. 1*A*). As determined by flow cytometric analysis of propidium iodide-stained cells, apoptosis failed to account for the reduced cell numbers in response to 5-Aza-CdR. Rather, growth inhibition was characterized by an increased proportion of cells in G_1 (data not shown). Treatment with 5-Aza-CdR depleted HT29 cells of soluble, nuclear DNA MeTase I as determined by Western analyses of nuclear protein

extracts (Fig. 1*B*). This depletion corresponded with the reexpression of a known methylation-silenced gene, p16 (8) (Fig. 1*C*). The kinetics of growth inhibition, depletion of DNA MeTase I, and the reactivation of p16 were consistent with the mechanistic properties of 5-Aza-CdR and verified our HT29 cell model system. Although the induction of p16 may contribute to the growth inhibition seen in response to 5-Aza-CdR (31), we hypothesized that the genomewide nature of 5-Aza-CdR-induced hypomethylation was likely to affect other growth inhibitory pathways.

5-Aza-CdR-Treatment Induces the Expression of IFN-Responsive Genes in HT29 Cells. To investigate the molecular mechanisms involved in 5-Aza-CdR-induced growth inhibition in HT29 cells, we constructed and utilized high-density cDNA microarrays to analyze gene expression changes coincident with 5-Aza-CdR treatment. Our array was composed of 4,608 randomly selected, minimally redundant cDNAs from the Unigene set (32). Labeled cDNA probes were prepared from vehicle-treated and 500 nM 5-Aza-CdR-treated HT29 cells 9 days after the initial drug exposure, a time that coincided with maximal growth inhibition (Fig. 1*A*). First-strand cDNAs were reverse-transcribed from mRNA samples in the presence of Cy-3dCTP (vehicle-treated) or Cy-5dCTP (5-Aza-CdR-treated). After labeling, the two probes were hybridized simultaneously to the microarray slide (Fig. 2*A*). Subsequent analysis revealed up-regulation of 19 genes by greater than 2 SD above the mean expression ratio for the entire gene set (Fig. 2*B*). We confirmed the induction of these genes with Northern analyses (Fig. 2*C*) and their identity by DNA sequencing (Table 1). We noted that 10 of 19 genes induced by 5-Aza-CdR were established IFN-response genes (Table 1) (27, 34–36). Because IFNs are established cell growth inhibitors (37–39), the stimulation of IFN-responsive genes in 5-Aza-CdR-treated HT29 cells presented an attractive hypothesis to explain the coincident growth inhibition (Fig. 1*A*).

To determine whether these 10 genes are regulated by IFN in HT29 cells, and to assess whether the other 9 genes also are responsive to IFN, we conducted microarray experiments on HT29 cells treated for 10, 24, or 96 hr with either IFN-α or IFN-γ (Fig. 3). Interestingly, each of the 19 genes regulated by 5-Aza-CdR were also induced by either IFN-α or IFN-γ (Table 1). Comparison of the induced genes revealed a significantly greater overlap between 5-Aza-CdR- and IFN-α-induced genes than between 5-Aza-CdR- and IFN-γ-induced genes (17/19 vs. 12/19 genes, respectively).

5-Aza-CdR Treatment Induces the Nuclear Accumulation and the Expression of STATs 1, 2, and 3. A simple explanation for the activation of IFN-responsive genes by 5-Aza-CdR is that the drug stimulated the synthesis and release of IFNs. Consistent with this possibility is the observation that the expression of IFN-γ can be regulated by DNA methylation (40–42). We, therefore, measured the mRNA levels for IFNs-α, -β, and -γ in HT29 after treatment with 500 nM 5-Aza-CdR. Reverse transcription–PCR analysis detected only IFN-α, and its mRNA level remained unchanged after 5-Aza-CdR treatment (Fig. 4*A*). We also were unable to detect any increase in IFN-α (or the presence of IFN-γ) by Western blot analysis of protein extracts from 5-Aza-CdR-treated HT29 cells at 2, 5, or 7 days after treatment (data not shown). These observations eliminated increased levels of IFNs as an explanation for the induction of IFN-responsive genes by 5-Aza-CdR. In addition, the transfer of medium harvested from 5-Aza-CdR-treated cells at 9 days after treatment onto control HT29 cell cultures did not inhibit cell growth. This indicates that the growth inhibition observed in 5-Aza-CdR-treated cells does not result from an increase in secreted IFN protein or of other growth inhibitory cytokines. These data caused us to investigate the IFN-signaling pathway to account for the induction of IFN-responsive genes by 5-Aza-CdR.

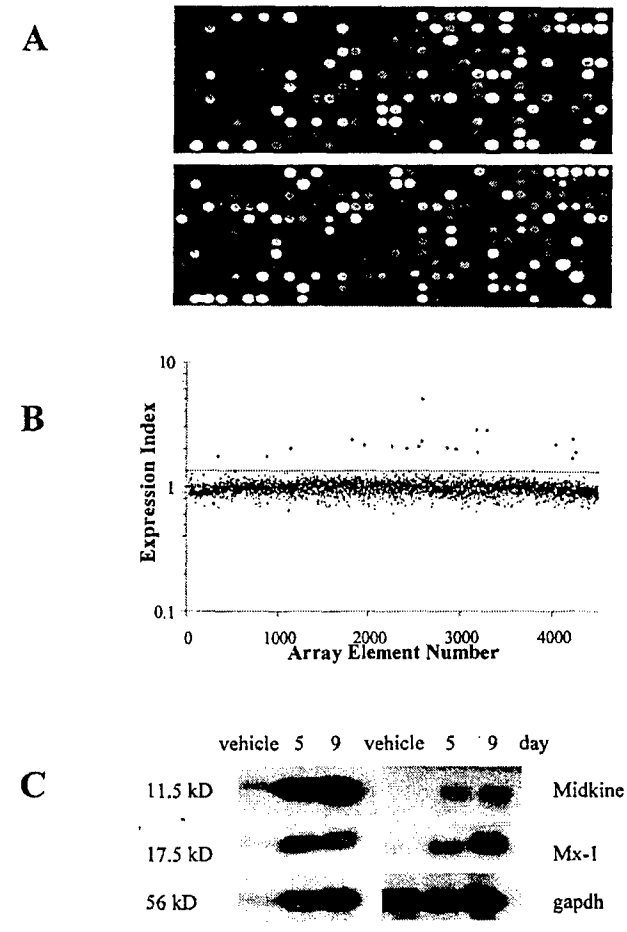


Fig. 2. Microarray analysis of gene expression changes in HT29 cells after 5-Aza-CdR treatment. (*A*) A cDNA microarray containing 4,608 target genes was constructed from a set of minimally redundant expressed sequence tags (ESTs). The microarray was hybridized with cDNAs prepared from vehicle-treated [Cy3-dCTP-labeled (green)] and 500 nM 5-Aza-CdR-treated HT29 cells [Cy5-dCTP (red)] 9 days after treatment. Two representative 12 × 32 gene grids (of 12) are displayed. (*B*) The fluorescent signal from the hybridized microarray slide was detected, quantified, and plotted as a ratio (Cy-5 signal/Cy-3 signal) for each array element. The average expression ratio for all genes on the array was normalized to 1.0 and had a SD of 0.177. The black line indicates a trend line 2 SD above the mean expression ratio for all genes on the microarray. The small, blue diamonds are genes below this cutoff; the large, red diamonds are genes above the cutoff. (*C*) Microarray expression data were confirmed by Northern blot analysis. Induction of five representative transcripts (see Table 1) 5 and 9 days after 5-Aza-CdR treatment is shown, along with glyceraldehyde-3-phosphate dehydrogenase (gapdh), an RNA-loading control. 11.5 kD, IFN-inducible protein 27; 17.5 kD, IFN-induced 17-kDa protein; 56 kD, IFN-induced protein 56.

Because STAT transcription factors are effectors of IFN signaling (43), we next examined whether 5-Aza-CdR treatment caused them to accumulate in the nuclei of HT29 cells. To address this, we performed Western blot analyses on fractionated HT29 cells by using antibodies specific for STATs 1, 2, 3, 4, 5, and 6. We observed a time-dependent increase of STATs 1, 2, and 3 in the nuclei of HT29 cells after treatment with 500 nM 5-Aza-CdR (Fig. 4*B*). In contrast, STATs 4, 5, and 6 did not accumulate in the nuclei after 5-Aza-CdR treatment (data not shown). We also noted an increase in the total cellular levels of STATs 1, 2, and 3 after 5-Aza-CdR treatment (Fig. 4*B*). This novel observation raised the possibility that inhibition of DNA MeTase induced the expression of STATs 1, 2, and 3.

Table 1. Genes up-regulated by 5-Aza-CdR treatment of HT29 cells*

5-Aza-CdR-induced gene†	Unigene number or image ID	Regulation by IFN-α‡	Regulation by IFN-γ‡
Human mRNA for Stac	Hs. 56045	+	+
IFN-induced protein 56	Hs. 20315	+	−
IFNα-inducible protein 27	Hs. 2867	+	+
IFN-induced 17-kDa protein	IID 149319	+	+
EST	Hs. 6166	+	−
EST	Hs. 165240	+	+
Myxovirus resistance gene 2	Hs. 926	+	−
Purinergic receptor P2Y5	Hs. 189999	+	+
EST	Hs. 109309	+	−
CpG island DNA fragment	No match§	+	+
TGF-β superfamily member MIC-1	Hs. 116577	+	+
IFN-induced protein IFI-6-16	Hs. 21205	+	−
EST	Hs. 47783	+	+
MHC class I	Hs. 77961	+	+
Midkine	Hs. 82045	−	+
Myxovirus resistance gene 1	Hs. 76391	+	−
2′-5′-Oligoadenylate synthetase 3	No match¶	+	+
Nuclear antigen SP100	Hs. 77617	+	−
IFN-inducible protein 10	Hs. 2248	−	+

Bold type indicates previously identified IFN-responsive genes.

*Genes induced in HT29 cells by treatment with 500 nM 5-Aza-CdR at day 9 that were up-regulated by greater than 2 SD above the mean expression ratio for all genes on the microarray (see Fig. 2*B*), listed in descending order.

†The identity of each gene was verified by DNA sequencing and BLAST analysis (33).

‡The expression of the genes induced by 5-Aza-CdR (column 1) was measured by microarray analysis after treatment of HT29 cells with 1×10^5 units/ml IFN-α2a or 5×10^2 units/ml IFN-γ for 10, 24, and 96 hr. Expression was scored as induced if the gene was up-regulated at any time point.

§Identical to accession numbers Z61029 and Z61030 (100% identity over 115 bases) in the nonredundant Genbank database.

¶Identical to accession number NM_006187 (100% identity over 366 bases) in the nonredundant Genbank database.

Because cDNAs corresponding to STATs 1, 2, and 3 were not on the microarray, we next performed Northern blot analyses on RNAs from 5-Aza-CdR-treated HT29 cells by using probes specific for STATs 1, 2, and 3. Fig. 4*C* illustrates the time-dependent up-regulation of STATs 1, 2, and 3 mRNA levels after 5-Aza-CdR treatment. This induction correlated temporally with growth inhibition in response to 5-Aza-CdR and implicates the transcriptional activation of STATs 1, 2, and 3 in the response of HT29 cells to 5-Aza-CdR. We also found that the STAT genes were expressed above control levels for at least 17 days (5 cell passages) after treatment with 5-Aza-CdR (data not shown).

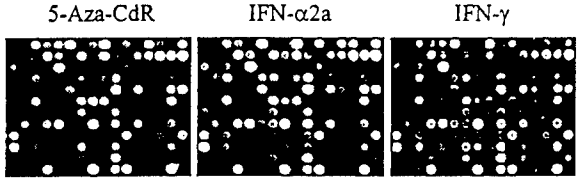


Fig. 3. Microarray expression profiling of 5-Aza-CdR and IFN-treated HT29 cells. HT29 cells were treated with 500 nM 5-Aza-CdR, 1×10^5 units/ml IFN-α2a, or 5×10^2 units/ml IFN-γ. RNA was harvested 9 days (5-Aza-CdR) or 4 days (96 hr) (IFN-α or -γ) after treatment and used to generate probes for microarray analysis. Shown in the figure is a representative section of the microarray after hybridization with Cy-5-labeled cDNAs from 5-Aza-CdR-, IFN-α-, or IFN-γ-treated cells and Cy-3-labeled cDNAs from control cells. Four genes up-regulated by 5-Aza-CdR treatment are on the displayed grid. They are IFN-α-inducible protein 6–16 (row 4, column 9), expressed sequence tag (EST) Hs.109309 (8, 7), EST Hs.165240 (9, 14), and human mRNA for Stac (9, 16).

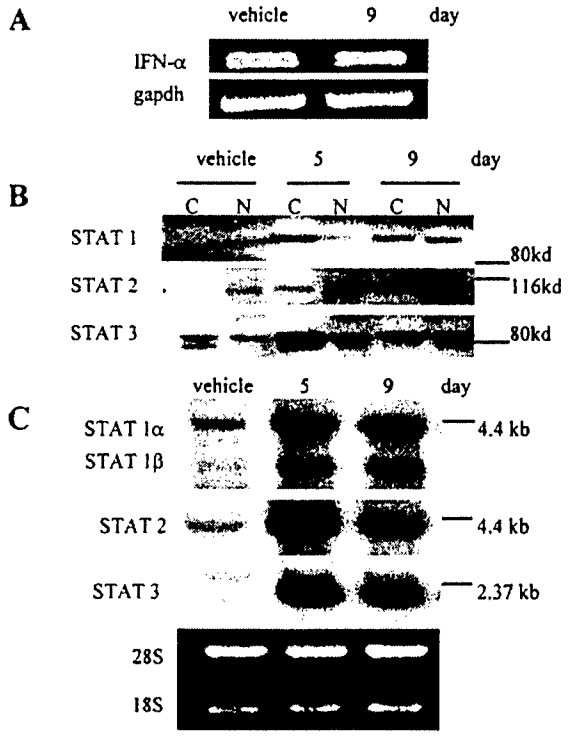


Fig. 4. 5-Aza-CdR treatment activates STATs 1, 2, and 3 in HT29 cells. (*A*) The expression level of IFN-α in HT29 cells before and after 500 nM 5-Aza-CdR treatment was measured by reverse transcription–PCR along with gapdh to confirm equivalent cDNA input. (*B*) STAT transcription factor levels were measured by Western blotting. Cytoplasmic (C) and nuclear (N) cell extracts were prepared from HT29 cells after treatment with vehicle or 500 nM 5-Aza-CdR. A poly(vinylidene difluoride) membrane harboring the protein extracts was probed sequentially with mAbs specific to STATs 1, 2, and 3. In each case, the antibodies recognized proteins of the appropriate molecular weight for each STAT. Molecular mass markers are indicated. (*C*) The expression of STAT 1, 2, and 3 genes was measured by Northern blotting. RNA was isolated from HT29 cells after treatment with vehicle or 500 nM 5-Aza-CdR. The locations of molecular mass markers are indicated. Ethidium bromide staining confirmed equal RNA loading (28S, 18S rRNAs).

5-Aza-CdR Treatment Sensitizes HT29 Cells to Exogenous IFN-α2a. The above data suggest that STATs 1, 2, and 3 limit the response of HT29 cells to IFNs. With this in mind, we hypothesized that 5-Aza-CdR treatment could potentiate the response of HT29 cells to IFN-α. To test this hypothesis, we exposed control and 5-Aza-CdR-treated HT29 cells to various concentrations of IFN-α2a and measured growth rates. We found that 5-Aza-CdR increased the responsiveness of HT29 cells to growth inhibition mediated by IFN-α2a (Fig. 5). This effect corresponded to at least a 5-fold increase in the potency (IC_{50} of 2×10^5 units IFN/ml for control cells vs. IC_{50} of 4×10^4 units IFN/ml for 5-Aza-CdR-treated cells) of IFN-α for inhibiting HT29 cell growth. It is important to note that the increased responsiveness was observed despite the high level of growth inhibition elicited by 5-Aza-CdR treatment alone (Fig. 1*A*).

Discussion

Transcriptional silencing of tumor-suppressor genes by CpG methylation may contribute to the development of human carcinomas. A model wherein methylation-induced gene silenc-

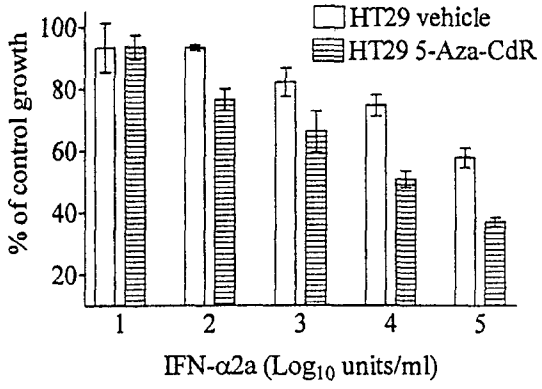


Fig. 5. 5-Aza-CdR treatment increases the responsiveness of HT29 cells to growth inhibition mediated by exogenous IFN-α2A. HT29 cells were treated with 500 nM 5-Aza-CdR or vehicle (PBS). Ten days after removal of the drug, triplicate wells were treated with a concentration curve of IFN-α2a. Four days later, cell proliferation was measured by using a Coulter counter. Percentage of control growth was calculated by dividing the mean cell count at each IFN concentration by the mean cell count of untreated control cells (either HT29 or 5-Aza-CdR-treated HT29 cells, respectively). Data are presented as mean ± 1 SD, (n = 3). Similar results were obtained in four independent experiments.

ing accompanies tumor development raises the potential for drug-induced reactivation of methylation-silenced tumor-suppressor genes as a therapeutic strategy. In this context, pharmacological inhibition of DNA MeTase by 5-Aza-CdR inhibits the growth of bladder, colon, and melanoma tumor cell lines, whereas control human fibroblasts are unaffected (31). Also, consistent with this model, a number of methylation-silenced tumor-suppressor genes have been identified by candidate gene approaches in tumor cells (3–5, 7, 11, 12, 44, 45). Among these, Bender *et al.* have demonstrated induction of p16 in a number of tumor cells that are growth-inhibited after 5-Aza-CdR treatment and that this induction correlates with the methylation status of the p16 promoter (31). However, it is reasonable to assume that the pharmacology of 5-Aza-CdR extends beyond p16-mediated growth arrest in that tumor cells in which p16 is not induced by 5-Aza-CdR are also growth-inhibited (31).

Our observation that 5-Aza-CdR inhibits HT29 cell growth parallels the results seen in other tumor cell lines (Fig. 1) (31) and validated them as a model system for microarray expression analysis. However, the results of our microarray analysis led to a new hypothesis for explaining the growth-inhibitory properties of 5-Aza-CdR in tumor cells *in vitro* and, perhaps, the efficacy of this compound *in vivo*. Our data indicate that STAT 1, 2, and 3 expression is induced by 5-Aza-CdR, that these proteins accumulate in the nucleus of 5-Aza-CdR-treated cells, and that these phenomenon parallel 5-Aza-CdR-induced growth inhibition. These data suggest that the presence of STAT proteins in tumor cells can dictate responsiveness to certain chemotherapeutics and raise the possibility that STATs 1, 2, and 3 are methylation-silenced tumor suppressors.

Our microarray approach started with an unbiased look at HT29 cell responses to 5-Aza-CdR and led us, indirectly, to the IFN-signaling pathway as a potential tumor-suppressive pathway. e saw that the genes responding most robustly to 5-Aza-CdR treatment in HT29 cells were also responsive to IFN treatment. This suggested the activation of the IFN-signaling pathway as a major cellular response to 5-Aza-CdR. The induction of IFN-responsive genes presents an attractive hypothesis for explaining 5-Aza-CdR-mediated growth inhibition in that IFNs are established growth-inhibitory cytokines (37, 39). However, it was unlikely that each of these IFN-responsive genes were regulated directly by promoter methylation. As an alternative, the microarray data pointed us to the up-regulation of STATs 1, 2, and 3, which are required to mediate the growth-inhibitory effects of IFN-γ (STAT 1) and IFN-α (STATs 1, 2, and 3) (43, 46, 47).

The up-regulation of STATs in response to 5-Aza-CdR may be explained in at least two ways. One explanation is that STAT genes are directly silenced by *de novo* methylation in tumor cells. In support of this model, the 5′ regions of STAT 1, 2, and 3 cDNAs contain likely CpG island targets for methylation (48) (see also GenBank accession no. L29277 for STAT 3). A second explanation is that induction of STATs 1, 2, and 3 by 5-Aza-CdR is the result of the epigenetic activation of another, upstream regulator of STAT expression. We do not believe it is likely that the stimulation of STATs and the IFN-induced gene set is due to nonspecific cellular toxicity or growth arrest because microarray experiments performed in our laboratory with agents such as TNF, TRAIL, FasL, and TGF-β have not revealed the induction of a similar gene set as that seen with 5-Aza-CdR and IFN-α and, to a lesser extent, with IFN-γ (data not shown). Whatever model accounts for the increased expression of STATs, it is unlikely that the simple up-regulation of these genes also results in their activation and nuclear accumulation. Rather, our data support a scenario in which endogenous IFN-α is responsible for activating STATs 1, 2, and 3. Several lines of evidence support this explanation. First, our analysis indicates the presence of IFN-α in control and 5-Aza-CdR-treated HT29 cells whereas IFN-β and -γ were undetectable under either condition. Second, our microarray analysis showed substantial overlap in genes induced by 5-Aza-CdR and those induced by the direct addition of IFN-α. Finally, our observation that STATs 1, 2, and 3 each accumulated in HT29 cell nuclei follows a number of studies demonstrating that IFN-α stimulation leads to activation of STAT1/2 or STAT1/3 heterodimers (43, 49, 50).

Our observation that 5-Aza-CdR stimulates expression of STATs 1, 2, and 3 holds important clinical implications. First, the expression of STAT 1 in certain metastatic melanoma and gastric adenocarcinoma cell lines is greatly depressed and correlates with a reduced level of responsiveness of these tumors to IFN-α (51–53). In the clinic, metastatic melanomas often fail to respond to IFN-α (54). Dampening of the IFN-response pathway by methylation silencing of STATs or other signaling components could account for lack of responsiveness of certain melanomas to IFN-α. Further, the activation of STATs by 5-Aza-CdR treatment raises the possibility that this drug could sensitize resistant tumor cells to IFN. As an initial test of this hypothesis, we examined the sensitivity of HT29 cells to the growth-inhibitory effects of IFN-α before and after treatment with 5-Aza-CdR. We saw that 5-Aza-CdR treatment increased the responsiveness of HT29 cells to IFN-α-mediated growth inhibition This result offers a plausible new line of investigation on the combination of 5-Aza-CdR and IFNs for the treatment of certain IFN-resistant tumors.

In conclusion, our work shows the value of microarray expression analyses in analyzing the mechanistic actions of pharmaceutical agents. Two previous studies have utilized microarrays to examine the specificity of drug actions in yeast. Gray *et al.* examined the transcriptional perturbations elicited by structural analogs of cyclin-dependent kinases inhibitors (28). In another study, Marton *et al.* compared the transcriptional profiles resulting from cyclosporin A and FK506 treatment of yeast mutant strains defective in calcineurin and immunophilin genes (29). These studies illustrate that microarrays can be used to examine drug-target specificity and potential secondary drug effects. We have extended these approaches by presenting a microarray-based evaluation of a clinically relevant compound in a human cell line. Although the explicit mechanistic basis for inhibition of DNA MeTase by 5-Aza-CdR is known, our study provides new, testable hypotheses that may explain the consequences of inhibiting DNA

methylation in a clinical setting. Further pharmacological studies that utilize microarrays are likely to reveal new lines of investigation, both *in vitro* and *in vivo*, that more focused experimental approaches may overlook.

We thank Ann Jones for cell culture expertise. We thank the DNA Sequencing Facility, DNA/Peptide Synthesis Facility, and Flow Cytometry Facilities at the University of Utah. This work was supported by grants from the Huntsman Cancer Foundation.

1. Nan, X., Ng, H. H., Johnson, C. A., Laherty, C. D., Turner, B. M., Eisenman, R. N. & Bird, A. (1998) *Nature (London)* **393,** 386–389.
2. Holliday, R. (1979) *Br. J. Cancer* **40,** 513–522.
3. Issa, J. P., Ottaviano, Y. L., Celano, P., Hamilton, S. R., Davidson, N. E. & Baylin, S. B. (1994) *Nat. Genet.* **7,** 536–540.
4. Ottaviano, Y. L., Issa, J. P., Parl, F. F., Smith, H. S., Baylin, S. B. & Davidson, N. E. (1994) *Cancer Res.* **54,** 2552–2555.
5. Yoshiura, K., Kanai, Y., Ochiai, A., Shimoyama, Y., Sugimura, T. & Hirohashi, S. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 7416–7419.
6. Herman, J. G., Umar, A., Polyak, K., Graff, J. R., Ahuja, N., Issa, J. P., Markowitz, S., Willson, J. K., Hamilton, S. R., Kinzler, K. W., *et al.* (1998) *Proc. Natl. Acad. Sci. USA* **95,** 6870–6875.
7. Gonzalez-Zulueta, M., Bender, C. M., Yang, A. S., Nguyen, T., Beart, R. W., Van Tornout, J. M. & Jones, P. A. (1995) *Cancer Res.* **55,** 4531–4535.
8. Bender, C. M., Zingg, J. M. & Jones, P. A. (1998) *Pharm. Res.* **15,** 175–187.
9. Bird, A. P. (1996) *Cancer Surv.* **28,** 87–101.
10. Laird, P. W., Jackson-Grusby, L., Fazeli, A., Dickinson, S. L., Jung, W. E., Li, E., Weinberg, R. A. & Jaenisch, R. (1995) *Cell* **81,** 197–205.
11. Veigl, M. L., Kasturi, L., Olechnowicz, J., Ma, A. H., Lutterbaugh, J. D., Periyasamy, S., Li, G. M., Drummond, J., Modrich, P. L., Sedwick, W. D., *et al.* (1998) *Proc. Natl. Acad. Sci. USA* **95,** 8698–8702.
12. Cote, S., Sinnett, D. & Momparler, R. L. (1998) *Anticancer Drugs* **9,** 743–750.
13. Momparler, R. L. (1985) *Pharmacol. Ther.* **30,** 287–299.
14. Juttermann, R., Li, E. & Jaenisch, R. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 11797–11801.
15. Ferguson, A. T., Vertino, P. M., Spitzner, J. R., Baylin, S. B., Muller, M. T. & Davidson, N. E. (1997) *J. Biol. Chem.* **272,** 32260–32266.
16. Kantarjian, H. M., O'Brien, S. M., Estey, E., Giralt, S., Beran, M., Rios, M. B., Keating, M., de Vos, D. & Talpaz, M. (1997) *Leukemia* **11,** Suppl 1, S35–S36.
17. Momparler, R. L., Rivard, G. E. & Gyger, M. (1985) *Pharmacol. Ther.* **30,** 277–286.
18. Momparler, R. L., Bouffard, D. Y., Momparler, L. F., Dionne, J., Belanger, K. & Ayoub, J. (1997) *Anticancer Drugs* **8,** 358–368.
19. Thibault, A., Figg, W. D., Bergan, R. C., Lush, R. M., Myers, C. E., Tompkins, A., Reed, E. & Samid, D. (1998) *Tumori* **84,** 87–89.
20. Jones, P. A. & Laird, P. W. (1999) *Nat. Genet.* **21,** 163–167.
21. Ushijima, T., Morimura, K., Hosoya, Y., Okonogi, H., Tatematsu, M., Sugimura, T. & Nagao, M. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 2284–2289.
22. Hatada, I., Sugama, T. & Mukai, T. (1993) *Nucleic Acids Res.* **21,** 5577–5582.
23. Gonzalgo, M. L., Liang, G., Spruck, C. H., III, Zingg, J. M., Rideout, W. M., III, & Jones, P. A. (1997) *Cancer Res.* **57,** 594–599.
24. Shiraishi, M., Chuu, Y. H. & Sekiya, T. (1999) *Proc. Natl. Acad. Sci. USA* **96,** 2913–2918.
25. Schena, M., Shalon, D., Davis, R. W. & Brown, P. O. (1995) *Science* **270,** 467–470.
26. Cho, R. J., Campbell, M. J., Winzeler, E. A., Steinmetz, L., Conway, A., Wodicka, L., Wolfsberg, T. G., Gabrielian, A. E., Landsman, D., Lockhart, D. J. & Davis, R. W. (1998) *Mol. Cell* **2,** 65–73.
27. Der, S. D., Zhou, A., Williams, B. R. & Silverman, R. H. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 15623–15628.
28. Gray, N. S., Wodicka, L., Thunnissen, A. M., Norman, T. C., Kwon, S., Espinoza, F. H., Morgan, D. O., Barnes, G., LeClerc, S., Meijer, L., *et al.* (1998) *Science* **281,** 533–538.
29. Marton, M. J., DeRisi, J. L., Bennett, H. A., Iyer, V. R., Meyer, M. R., Roberts, C. J., Stoughton, R., Burchard, J., Slade, D., Dai, H., *et al.* (1998) *Nat. Med.* **4,** 1293–1301.
30. Wang, P., Wu, P., Siegel, M. I., Egan, R. W. & Billah, M. M. (1995) *J. Biol. Chem.* **270,** 9558–9563.
31. Bender, C. M., Pao, M. M. & Jones, P. A. (1998) *Cancer Res.* **58,** 95–101.
32. Miller, G. S. & Fuchs, R. (1997) *Comput. Appl. Biosci.* **13,** 81–87.
33. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997) *Nucleic Acids Res.* **25,** 3389–3402.
34. Stadler, M., Chelbi-Alix, M. K., Koken, M. H., Venturini, L., Lee, C., Saib, A., Quignon, F., Pelicano, L., Guillemin, M. C., Schindler, C., *et al.* (1995) *Oncogene* **11,** 2565–2573.
35. Rasmussen, U. B., Wolf, C., Mattei, M. G., Chenard, M. P., Bellocq, J. P., Chambon, P., Rio, M. C. & Basset, P. (1993) *Cancer Res.* **53,** 4096–4101.
36. Luster, A. D., Jhanwar, S. C., Chaganti, R. S., Kersey, J. H. & Ravetch, J. V. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 2868–2871.
37. Pfeffer, L. M., Dinarello, C. A., Herberman, R. B., Williams, B. R., Borden, E. C., Bordens, R., Walter, M. R., Nagabhushan, T. L., Trotta, P. P. & Pestka, S. (1998) *Cancer Res.* **58,** 2489–2499.
38. Stark, G. R., Kerr, I. M., Williams, B. R., Silverman, R. H. & Schreiber, R. D. (1998) *Annu. Rev. Biochem.* **67,** 227–264.
39. Nathan, C. (1992) in *Inflammation: Basic Principles and Clinical Correlates*, eds. Gallin, J. I. & Goldstein, I. M. (Raven, New York), 2nd Ed., pp. 265–289.
40. Fukunaga, R., Matsuyama, M., Okamura, H., Nagata, K., Nagata, S. & Sokawa, Y. (1986) *Nucleic Acids Res.* **14,** 4421–4436.
41. Young, H. A., Ghosh, P., Ye, J., Lederer, J., Lichtman, A., Gerard, J. R., Penix, L., Wilson, C. B., Melvin, A. J., McGurn, M. E., *et al.* (1994) *J. Immunol.* **153,** 3603–3610.
42. Mikovits, J. A., Young, H. A., Vertino, P., Issa, J. P., Pitha, P. M., Turcoski-Corrales, S., Taub, D. D., Petrow, C. L., Baylin, S. B. & Ruscetti, F. W. (1998) *Mol. Cell. Biol.* **18,** 5166–5177.
43. Darnell, J. E., Jr. (1997) *Science* **277,** 1630–1635.
44. Belinsky, S. A., Nikula, K. J., Palmisano, W. A., Michels, R., Saccomanno, G., Gabrielson, E., Baylin, S. B. & Herman, J. G. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 11891–11896.
45. Lapidus, R. G., Ferguson, A. T., Ottaviano, Y. L., Parl, F. F., Smith, H. S., Weitzman, S. A., Baylin, S. B., Issa, J. P. J. & Davidson, N. E. (1996) *Clin. Cancer Res.* **2,** 805–810.
46. Bromberg, J. F., Horvath, C. M., Wen, Z., Schreiber, R. D. & Darnell, J. E., Jr. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 7673–7678.
47. Ihle, J. N., Thierfelder, W., Teglund, S., Stravapodis, D., Wang, D., Feng, J. & Parganas, E. (1998) *Ann. N. Y. Acad. Sci.* **865,** 1–9.
48. Yan, R., Qureshi, S., Zhong, Z., Wen, Z. & Darnell, J. E., Jr. (1995) *Nucleic Acids Res.* **23,** 459–463.
49. Yang, C. H., Murti, A. & Pfeffer, L. M. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 5568–5572.
50. Stancato, L. F., David, M., Carter-Su, C., Larner, A. C. & Pratt, W. B. (1996) *J. Biol. Chem.* **271,** 4134–4137.
51. Abril, E., Real, L. M., Serrano, A., Jimenez, P., Garcia, A., Canton, J., Trigo, I., Garrido, F. & Ruiz-Cabello, F. (1998) *Cancer Immunol. Immunother.* **47,** 113–120.
52. Sun, W. H., Pabon, C., Alsayed, Y., Huang, P. P., Jandeska, S., Uddin, S., Platanias, L. C. & Rosen, S. T. (1998) *Blood* **91,** 570–576.
53. Wong, L. H., Krauer, K. G., Hatzinisiriou, I., Estcourt, M. J., Hersey, P., Tam, N. D., Edmondson, S., Devenish, R. J. & Ralph, S. J. (1997) *J. Biol. Chem.* **272,** 28779–28785.
54. Punt, C. J. (1998) *Melanoma Res.* **8,** 95–104.

[CANCER RESEARCH 59, 6042–6045, December 15, 1999]

Advances in Brief

E7-transduced Human Breast Epithelial Cells Show Partial Differentiation in Three-dimensional Culture[1]

Kimberly M. Spancake,[2] Christine B. Anderson, Valerie M. Weaver, Norisada Matsunami, Mina J. Bissell, and Raymond L. White

Huntsman Cancer Institute, University of Utah, Salt Lake City, Utah 84112 [K. M. S., C. B. A., N. M., R. L. W.], and Lawrence Berkeley National Laboratory, Berkeley, California 94720 [V. M. W., M. J. B.]

Abstract

Disruption of the retinoblastoma (RB) tumor suppressor pathway is a common and important event in breast carcinogenesis. To examine the role of the retinoblastoma protein (pRB) in this process, we created human mammary epithelial cells (HMEC) deficient for pRB by infecting primary outgrowth from breast organoids with the human papillomavirus type 16 (HPV16) *E7* gene. HPV16 E7 binds to and inactivates pRB and also causes a significant down-regulation of the protein. Culturing normal HMEC in a reconstituted basement membrane (rBM) provides a correct environment and signaling cues for the formation of differentiated, acini-like structures. When cultured in this rBM, HMEC+E7 were found to respond morphologically as normal HMEC and form acinar structures. In contrast to normal HMEC, many of the cells within the HMEC+E7 structures were not growth arrested, as determined by a 5-bromo-2′-deoxyuridine incorporation assay. pRB deficiency did not affect polarization of these structures, as indicated by the normal localization of the cell-cell adhesion marker E-cadherin and the basal deposition of a collagen IV membrane. However, in HMEC+E7 acini, we were unable to detect by immunofluorescence microscopy the milk protein lactoferrin or cytokeratin 19, both markers of differentiation expressed in the normal HMEC structures. These data suggest that loss of *RB in vivo* would compromise differentiation, predisposing these cells to future tumor-promoting actions.

Introduction

The tumor suppressor gene *RB*[3] can play a significant role in breast carcinogenesis. *RB* has been shown to be inactivated in 19% of human breast tumors and 25% of human breast carcinoma cell lines (1, 2). Other members of the RB regulatory network have also been reported as having aberrant expression in a significant number of breast carcinomas. For example, cyclin D1, which through regulation of the cyclin-dependent kinase Cdk4 phosphorylates and inactivates pRB, is overexpressed in 35% of breast tumors (3). The p16 protein, also a regulator of pRB activity through inhibition of Cdk4, is absent in 40% of breast tumors, as determined by immunohistochemical analysis (4). Additionally, restoration of pRB expression in human breast cancer cell lines lacking functional pRB has been shown to cause a reduction in the cells' tumorigenicity (5). These data suggest that disruption of the RB regulatory network is a common and important step in breast carcinogenesis.

In addition to the cellular inactivation of pRB by cyclins and cyclin-dependent kinases, several DNA tumor virus proteins abrogate pRB activity by binding to and functionally sequestering pRB. The HPV16 E7 protein not only inactivates pRB function by binding but also enhances degradation of pRB through ubiquitin-dependent proteolysis (6).

HPV16 E7 was shown previously to immortalize early passage HMEC in culture (7), suggesting a critical role in growth regulation for pRB *in vivo*. Disruption of such fundamental growth control mechanisms should generate cell populations predisposed for transformation. To create a pRB-deficient cell line, we infected primary breast epithelial cell outgrowth with a retrovirus expressing the HPV16 *E7* gene (HMEC+E7). In this study, only early passage, G418-resistant, precrisis HMEC+E7 cells were examined. To characterize the impact of this pRB deficiency, we used a sensitive three-dimensional culture assay shown previously to support the formation of spheroid structures that resemble breast acini (8). Culturing human, luminal breast epithelial cells in a rBM provides an environment that supports not only organogenesis but also formation of endogenous BM necessary for differentiation (9). This culture system can also be used to distinguish between normal and cancer cells because of the inability of the transformed cells to organize into acinar structures (10).

We examined the effect of pRB deficiency on structure formation, cell-cell interactions, and differentiation by culturing the cells in a rBM. Our data demonstrate that even in the absence of functional pRB, human breast epithelial cells organize and form morphologically typical acini-like structures. Furthermore, in these pRB-deficient structures, proper localization of E-cadherin and deposition of a collagen IV BM, both markers of organogenesis, were found. However, unlike normal HMEC, the structures formed by HMEC+E7 do not exit the cell cycle and are deficient for the expression of cytokeratin 19 and lactoferrin, two proteins associated with the differentiation of luminal breast epithelial cells. These data suggest that pRB is not necessary for the transmission of signals from the extracellular matrix that convey information necessary for structure formation, but pRB does appear to be necessary for cell cycle withdrawal and expression of some proteins normally present in the differentiated luminal epithelial cell. Our data suggest that the initial loss of *RB in vivo* would not predispose breast epithelial cells to an apoptotic or immediate transformed fate, but rather loss of pRB activity would contribute to a less differentiated cellular state that, in turn, may eventually lead to a malignant phenotype.

Materials and Methods

Cell Culture and Retroviral Infection. Surgical discard material from reduction mammoplasties was minced with opposing scalpels, placed in digestion buffer, and incubated in spinner flasks at 37°C until stroma dissolved (~3–5 h). Digestion buffer contained 1 unit/ml Collagenase D (Roche Molecular Biochemicals, Indianapolis, IN), 2.4 units/ml dispase (Roche Molecular Biochemicals), and 6.25 units/ml DNase (Sigma Chemical Co., St. Louis, MO) in Dulbecco's PBS. The digested material plus 10% FCS was centrifuged at 800 rpm for 10 min. The resulting pellet was resuspended in wash buffer (11), and the organoids were

Received 8/26/99; accepted 11/1/99.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] This work was supported in part by the United States Army Medical Research and Materiel Command Grant DAMD-17-94-J-4154 and the Huntsman Cancer Institute.

[2] To whom requests for reprints should be addressed, at Huntsman Cancer Institute, University of Utah, 2000 Circle of Hope, Building 555, Room 567B, Salt Lake City, Utah 84112. Phone: (801) 581-6401; Fax: (801) 585-0374; E-mail: kimberly.spancake@hci.utah.edu.

[3] The abbreviations used are: RB, retinoblastoma; pRB, retinoblastoma protein; HMEC, human mammary epithelial cells; HPV16, human papillomavirus type 16; rBM, reconstituted basement membrane; DAPI, 4,6-diamidino-2-phenylindole; BrdUrd, 5-bromo-2′-deoxyuridine.

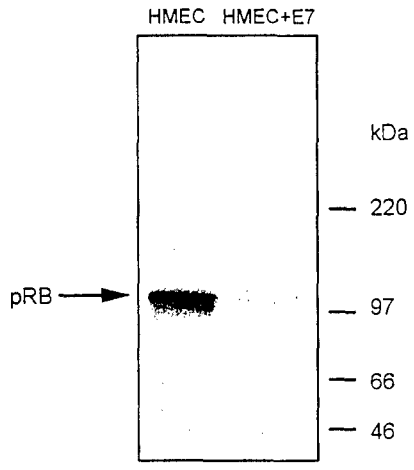


Fig. 1. Western blot analysis of pRB in normal and HPV16 E7-expressing breast epithelial cells. Protein lysates from HMEC and HMEC+E7 were analyzed by SDS-PAGE and immunoblotting with a monoclonal anti-pRB antibody. pRB is present in the normal HMEC, but the level of expression is reduced in HMEC+E7.

separated from stromal and blood cells by sequential sedimentation at 1 × g. Organoids were cultured immediately or frozen in DMEM: Nutrient Mixture F-12 (Ham) 1:1 (Life Technologies, Inc., Gaithersburg, MD), 10% FCS, and 10% DMSO. Organoids were cultured on plastic in CDM3 culture medium (12) for primary breast epithelial cell outgrowth and subculture. Primary epithelial outgrowth was infected with an LXSN retroviral construct containing the HPV16 *E7* gene (LXSN16E7; Ref. 13). HMEC were incubated with LXSN16E7 in CDM3 plus 4 μg/ml Polybrene (Sigma) for 24 h. Viral supernatant was aspirated, and HMEC were cultured in virus-free CDM3 for 48 h prior to selection in CDM3 containing 50 μg/ml Geneticin (Life Technologies, Inc.). Early-passage HMEC (either passage one or two) and HMEC containing LXSN16E7 (HMEC+E7) were cultured in a rBM, Matrigel (Becton Dickinson, Bedford, MA), as described previously (10). Briefly, 2.5×10^5 HMEC were resuspended as single cells in 300 μl of 10 mg/ml Matrigel per well and plated onto Nunc four-well multidishes coated with 100 μl of Matrigel. Matrigel cultures were overlaid with 500 μl of CDM3.

Western Blot Analysis. HMEC and HMEC+E7 cultured in polystyrene flasks were scraped in lysis buffer (PBS containing 0.1% Triton X-100, 0.1% NP40, 0.2 mg/ml Pefablock, 0.01 mg/ml aprotinin, 0.01 mg/ml pepstatin, 0.01 mg/ml leupeptin, 20 mM sodium fluoride, and 1 mM sodium orthovanadate) and sonicated for 30 s on ice. HMEC and HMEC+E7 cultured in rBM for 10 days were liberated from the Matrigel by incubation with dispase (Becton Dickinson) for 1 h at 37°C. The acini were then washed with PBS plus 5 mM EDTA, resuspended in lysis buffer, homogenized for 10 s, and sonicated for 30 s on ice. Protein lysates from cells cultured on plastic or in Matrigel were boiled in SDS-PAGE sample buffer for 3 min, and 50 μg of total protein were resolved on 4–12% gradient acrylamide Tris-Glycine gels (Novex, San Diego, CA) by SDS-PAGE. Proteins were transferred to nitrocellulose membrane (Schleicher & Schuell, Dassel, Germany). Immunoblot analysis was performed as described previously (14). Antibodies were used in the following concentrations: RB (mouse IgG1, G3–245) 1:400 (PharMingen, San Diego, CA), cytokeratin 19 (mouse IgG1, RCK 108) 1:100 (DAKO), and horseradish peroxidase-goat antimouse (Sigma) 1:30,000.

Immunofluorescence. Ten-day HMEC and HMEC+E7 rBM cultures were fixed, frozen, and cut into 5-μm sections as described previously (15). Antibodies were used in the following concentrations: E-cadherin (mouse IgG2a, 36) 1:100 (Transduction Laboratories, Lexington, KY), collagen IV (mouse IgG1, CIV 22) 1:50 (DAKO, Carpinteria, CA), cytokeratin 18 (mouse IgG1, CY-90) 1:800 (Sigma), lactoferrin (rabbit sera) 1:50 (Zymed, San Francisco, CA), cytokeratin 19 (mouse IgG1, RCK 108) 1:100 (DAKO), goat antimouse IgG1-FITC 1:200 (Southern Biotechnology Associates, Birmingham, AL), Alexa 488 goat antimouse IgG (H+L), F(ab')$_2$ fragment conjugate 1:1000 (Molecular Probes, Eugene, OR), and goat antirabbit-Texas Red 1:200 (Accurate Chemical and Scientific, Westbury, NY). Nuclei were counterstained with 100 ng/ml DAPI (Sigma) or TO-PRO-3 1:750 (Molecular Probes).

Proliferation Assay. rBM cultures were prepared as above with the following exceptions. Ten μM BrdUrd was added to the culture medium 12 h prior to freezing the culture without fixative. Cryosections were fixed in 2% paraformaldehyde in PBS for 10 min, followed by a 30-min incubation in 2 M HCl, all performed at room temperature. Staining proceeded as above using anti-BrdUrd (mouse IgG1, BMC9318) 1:10 and sheep antimouse immunoglobulin-FITC 1:10 (Boehringer Mannheim, Indianapolis, IN). Slides were scored visually (~200 cells/experiment) for BrdUrd-labeled nuclei, and indices were calculated by expressing this number as a percentage of the total nuclei scored (10).

Results

HMEC expressing HPV16 E7 Show a Decrease in pRB Expression. Primary epithelial outgrowth from normal human organoids was infected with a retroviral construct expressing the HPV16 *E7* gene. HMEC+E7 were selected by culturing the cells in CDM3 growth medium containing Geneticin. To determine the effect of HPV16 E7 expression on pRB, HMEC and HMEC+E7 were examined by Western immunoblot analysis. Fig. 1 shows the expression level of pRB in the normal, parental HMEC when cultured on plastic. In contrast, HMEC+E7 showed a significant decrease in pRB. This down-regulation of pRB is in agreement with previous results demonstrating the ubiquitin-dependent proteolysis of pRB in the presence of HPV16 E7 (6).

HMEC+E7 Form Acinar Structures When Cultured in Extracellular Matrix But Are Not Growth Arrested. Both normal HMEC and HMEC+E7 were cultured in rBM. By 10 days in culture,

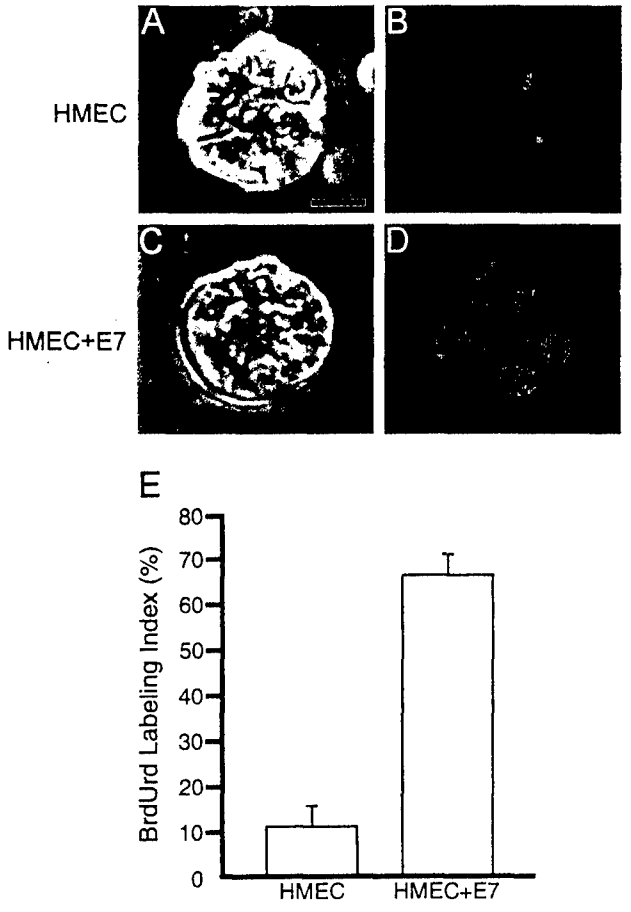


Fig. 2. HMEC+E7 cultured in extracellular matrix form acinar structures that are not growth arrested. Breast cells were cultured in a rBM for 10 days. *A* and *C*, phase contrast images of representative HMEC (*A*) and HMEC+E7 (*C*) structures. *B* and *D*, frozen sections of HMEC (*B*) and HMEC+E7 (*D*) acini stained with DAPI. HMEC+E7 cultured in rBM formed structures morphologically similar to HMEC, although the nuclei appear larger and slightly less organized. Percentage of BrdUrd labeling in HMEC and HMEC+E7 structures is expressed as the BrdUrd labeling index from three separate experiments (*E*; *bars*, SD). Approximately 90% of HMEC were growth arrested as compared with HMEC+E7 that continued to synthesize DNA. *Bar*, 20 μm.

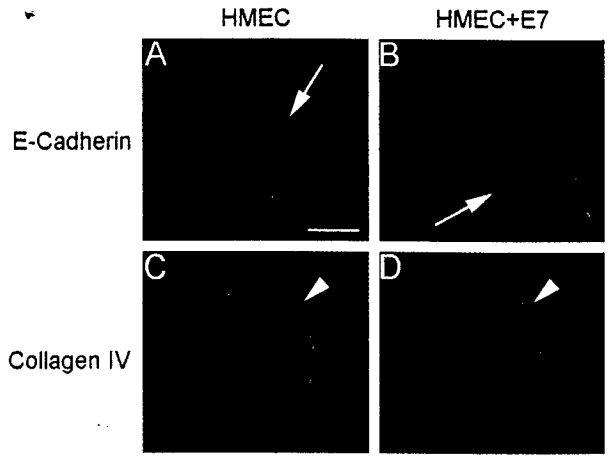


Fig. 3. Localization of E-cadherin and collagen IV in HMEC and HMEC+E7 using immunofluorescence microscopy. *A* and *B*, frozen sections of HMEC and HMEC+E7 acini, respectively, were stained with a monoclonal anti-E-cadherin antibody, Alexa 488 goat antimouse IgG (H+L), $F(ab')_2$ fragment conjugate, and the nuclei were counterstained with TO-PRO-3. Confocal fluorescence microscopy showed E-cadherin localization at points of cell-cell contact (*A* and *B*, *arrows*). *C* and *D*, frozen sections of HMEC and HMEC+E7 acini, respectively, were stained with a monoclonal anti-collagen IV antibody and goat antimouse IgG1-FITC. Collagen IV was localized in a continuous BM (*arrowheads*) at the basal surface of HMEC (*C*) and HMEC+E7 (*D*) structures. *Bar*, 20 μm.

normal HMEC organized into differentiated acini ~55 μm in diameter (Fig. 2*A*). A 5-μm frozen section of a representative structure stained with DAPI revealed a ring of basally located nuclei (Fig. 2*B*). HMEC+E7 also formed morphologically normal-appearing acinar structures when cultured in rBM (Fig. 2*C*). The HMEC+E7 acini were similar in size to the parental structures, although DAPI staining revealed somewhat larger and slightly less organized nuclei (Fig. 2*D*). Additionally, HMEC acini had growth arrested, as determined by minor BrdUrd incorporation similar to levels reported previously (15); however, a significant increase in BrdUrd labeling was observed in HMEC+E7 (Fig. 2*E*).

Normal Localization of E-cadherin and Type IV Collagen in HMEC+E7 Acinar Structures. HMEC+E7 were examined for the expression of proteins shown previously to have distinctive localization patterns characteristic of organogenesis in normal HMEC acini (15). HMEC and HMEC+E7 rBM cultures were fixed, frozen, cut into 5-μm sections, and stained with the indicated primary antibodies and corresponding secondary antibodies. In normal HMEC acinar structures, the cell-cell adhesion protein E-cadherin was localized to points of cell-cell contact (Fig. 3*A*). This same pattern of E-cadherin expression at points of cell-cell contact was present in the HMEC+E7 structures (Fig. 3*B*). Collagen IV was basally deposited in a continuous BM surrounding the normal HMEC structures when cultured in rBM (Fig. 3*C*). HMEC+E7 acini also deposited a basal collagen IV-containing BM (Fig. 3*D*).

Lactoferrin and Cytokeratin 19 Are Not Expressed in Structurally Differentiated HMEC+E7. The proteins lactoferrin and cytokeratin 19 are markers for differentiation in luminal breast epithelial cells (16–18). HMEC and HMEC+E7 were prepared as described above and examined for the expression of lactoferrin and cytokeratin 19. HMEC and HMEC+E7 acinar structures were identified by the presence of the luminal marker cytokeratin 18 (Fig. 4*A* and *B*, respectively) and the absence of the myoepithelial marker cytokeratin 14 (data not shown). Lactoferrin is an iron-binding milk protein that was expressed in the normal HMEC acini (Fig. 4*C*). In contrast, immunofluorescence analysis of HMEC+E7 structures demonstrated the absence of lactoferrin (Fig. 4*D*). As expected, structures derived from normal HMEC expressed the luminal differentiation marker cytokeratin 19 (Fig. 4*E*). HMEC+E7 acini, however, were deficient for the expression of cytokeratin 19 (Fig. 4*F*). These immunofluorescence data were confirmed by Western immunoblot analysis of HMEC and HMEC+E7 acinar structures, which showed barely detectable levels of cytokeratin 19 in HMEC+E7 (Fig. 4*G*).

Discussion

HMEC+E7 are deficient for pRB, a key protein governing cell cycle regulation and differentiation; thus, formation of even a partially "differentiated" structure when cultured in rBM was unexpected. Further investigation showed that although these pRB-deficient cells were capable of organizing into morphologically normal acini, the nuclei of these structures appeared to be larger and slightly less

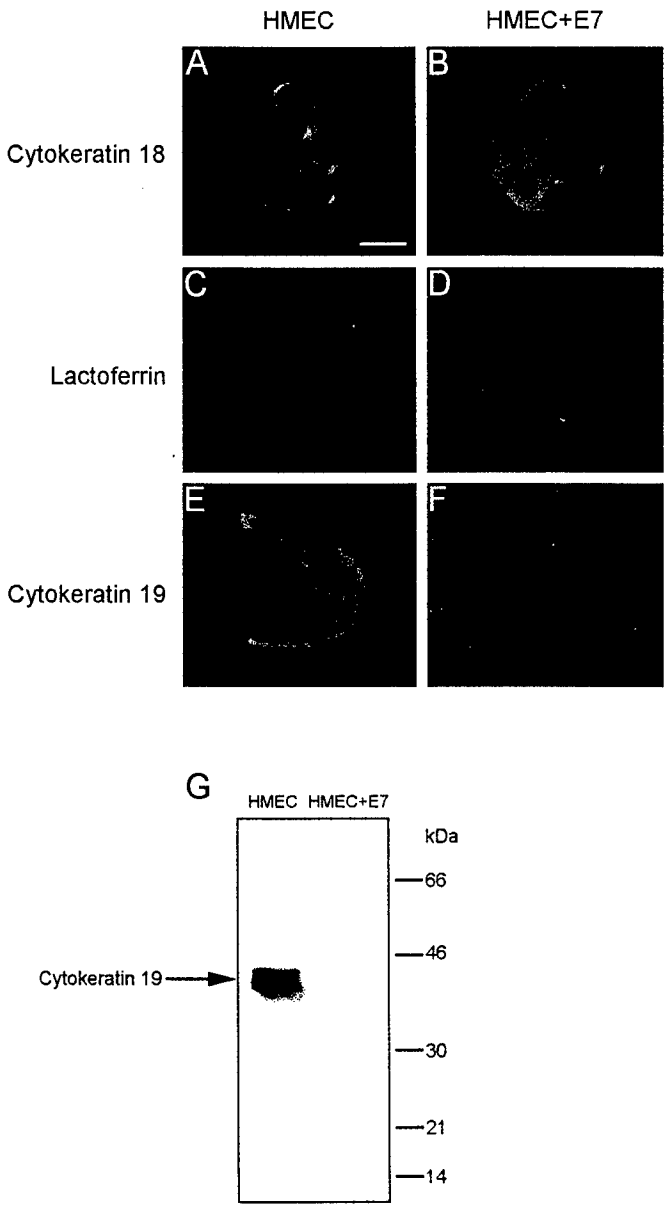


Fig. 4. Expression of cytokeratin 18, lactoferrin, and cytokeratin 19 in HMEC and HMEC+E7 acinar structures using immunofluorescence microscopy and Western immunoblot analysis. *A* and *B*, frozen sections of HMEC and HMEC+E7, respectively, were stained with a monoclonal cytokeratin 18 antibody and goat antimouse IgG1-FITC to confirm that structures were of luminal origin. Frozen sections of HMEC (*C* and *E*) and HMEC+E7 (*D* and *F*) were stained for lactoferrin (*C* and *D*), cytokeratin 19 (*E* and *F*), and the nuclei were visualized with a DAPI counterstain (*C* and *D*). Protein lysates from HMEC and HMEC+E7 acinar structures were analyzed by SDS-PAGE and immunoblotting with a monoclonal anti-cytokeratin 19 antibody (*G*). HMEC+E7 acinar structures were deficient for both lactoferrin and cytokeratin 19 expression. *Bar*, 20 μm.

organized. Additionally, HMEC+E7 acini contained many cells that had not growth arrested, unlike the normal HMEC structures. Immunofluorescence analysis of the HMEC+E7 structures revealed normal patterns of E-cadherin and collagen IV localization, indicating that the HMEC+E7 were forming normal cell-cell contacts and were properly polarized. HMEC+E7 acini expressing the luminal marker cytokeratin 18 lacked proteins typically expressed in the differentiated luminal breast epithelial cell. Neither the milk protein lactoferrin nor the luminal differentiation marker cytokeratin 19 were found in the HMEC+E7 acinar structures.

pRB has a significant role in the differentiation of many different cell types, including muscle, neuronal, and erythroid cells (19). Typically, an early event in differentiation is the activation of pRB by dephosphorylation, allowing the cell to arrest in G_0-G_1. During muscle cell differentiation, hypophosphorylated pRB associates with the transcription factors MyoD and myogenin, preventing phosphorylation of pRB and the consequent reemergence into the cell cycle (20). pRB has not been shown previously to have a role in directing the differentiation of breast epithelial cells. Because pRB is down-regulated in HMEC+E7, our data suggest that pRB is not necessary during the organization and structure formation of HMEC when cultured in a three-dimensional rBM. However, the absence of pRB does impact the expression of some proteins normally associated with the differentiated phenotype. Interestingly, a similar observation was made during MyoD-induced skeletal muscle differentiation in which cells showed attenuated expression of a late differentiation marker, myosin heavy chain (21). The deficiency of differentiation markers in the HMEC+E7 structures could be attributable to the role of pRB in transcriptional regulation outside its classically described interaction with E2F (reviewed in Ref. 19). Recently, pRB was shown to interact directly with the transcription factor SP1 and increase transcription of the *dihydrofolate reductase* gene via the SP1 binding site (22). Sequence analysis of human lactoferrin and mouse cytokeratin 19 has revealed SP1 sites in the promoter regions of both these genes (23, 24). Taken together, these data suggest a possible role for pRB in transcriptionally regulating lactoferrin and cytokeratin 19, a control that is abolished by HPV16 E7.

HPV16 E7 protein can also interact with the pRB family members p107 and p130. However, neither p107 nor p130 have been shown to be targeted for ubiquitin-dependent proteolysis by HPV16 E7, as has been established for pRB (25). Additionally, to date no human tumors have been identified that contain mutations in p107 or p130 (reviewed in Ref. 26). Consequently, the roles of p107 and p130 in transformation remain unclear, and inactivation of these proteins could also be contributing to the phenotype observed in the HMEC+E7 structures. Our data demonstrate that breast epithelial cells with down-regulated pRB retain the ability to respond to structure-forming signaling cues from a rBM; however, these acinar structures are not growth arrested and do not express some of the proteins normally associated with differentiation. These data suggest that mutation of *RB* alone *in vivo* in human breast epithelial cells would not cause transformation but rather create a lesser differentiated cellular state. This modified phenotype, in conjunction with additional mutations, growth factor and hormonal modulations, and extracellular matrix perturbations, would likely result in malignancy.

Finally, it should be noted that the retinoblastoma protein has a complex set of functions within cells, and finding conditions that allow separation of these functions has been challenging. Our observations indicate that culture in a three-dimensional matrix affords a new operational definition of pRB activities that distinguishes those functions involved in polarity and initiation of differentiation from those functions that define a fully differentiated cell.

Acknowledgments

We are grateful to Dr. Kristi L. Neufeld for useful comments on the manuscript and Diana Lim for assistance with figure preparation.

References

1. Varley, J. M., Armour, J., Swallow, J. E., Jeffreys, A. J., Ponder, B. A., T'Ang, A., Fung, Y. K., Brammar, W. J., and Walker, R. A. The retinoblastoma gene is frequently altered, leading to loss of expression in primary breast tumours. Oncogene, *4:* 725–729, 1989.
2. T'Ang, A., Varley, J. M., Chakraborty, S., Murphree, A. L., and Fung, Y. K. Structural rearrangement of the retinoblastoma gene in human breast carcinoma. Science (Washington DC), *242:* 263–266, 1988.
3. Gillett, C., Fantl, V., Smith, R., Fisher, C., Bartek, J., Dickson, C., Barnes, D., and Peters, G. Amplification and overexpression of cyclin D1 in breast cancer detected by immunohistochemical staining. Cancer Res., *54:* 1812–1817, 1994.
4. Geradts, J., and Wilson, P. A. High frequency of aberrant p16(INK4A) expression in human breast cancer. Am. J. Pathol., *149:* 15–20, 1996.
5. Wang, N. P., To, H., Lee, W. H., and Lee, E. Y. Tumor suppressor activity of *RB* and *p53* genes in human breast carcinoma cells. Oncogene, *8:* 279–288, 1993.
6. Boyer, S. N., Wazer, D. E., and Band, V. E7 protein of human papilloma virus-16 induces degradation of retinoblastoma protein through the ubiquitin-proteasome pathway. Cancer Res., *56:* 4620–4624, 1996.
7. Wazer, D. E., Liu, X. L., Chu, Q., Gao, Q., and Band, V. Immortalization of distinct human mammary epithelial cell types by human papilloma virus 16 E6 or E7. Proc. Natl. Acad. Sci. USA, *92:* 3687–3691, 1995.
8. Li, M. L., Aggeler, J., Farson, D. A., Hatier, C., Hassell, J., and Bissell, M. J. Influence of a reconstituted basement membrane and its components on casein gene expression and secretion in mouse mammary epithelial cells. Proc. Natl. Acad. Sci. USA, *84:* 136–140, 1987.
9. Bissell, M. J., Weaver, V. M., Lelievre, S. A., Wang, F., Petersen, O. W., and Schmeichel, K. L. Tissue structure, nuclear organization, and gene expression in normal and malignant breast. Cancer Res., *59:* 1757s–1764s, 1999.
10. Petersen, O. W., Ronnov-Jessen, L., Howlett, A. R., and Bissell, M. J. Interaction with basement membrane serves to rapidly distinguish growth and differentiation pattern of normal and malignant human breast epithelial cells. Proc. Natl. Acad. Sci. USA, *89:* 9064–9068, 1992.
11. Taylor-Papadimitriou, J., and Stampfer, M. R. Culture of human mammary epithelial cells. *In:* R. I. Freshney (ed.), Culture of Epithelial Cells, pp. 108–133. New York: John Wiley & Sons, Inc., 1992.
12. Petersen, O. W., and van Deurs, B. Preservation of defined phenotypic traits in short-term cultured human breast carcinoma derived epithelial cells. Cancer Res., *47:* 856–866, 1987.
13. Halbert, C. L., Demers, G. W., and Galloway, D. A. The *E7* gene of human papillomavirus type 16 is sufficient for immortalization of human epithelial cells. J. Virol., *65:* 473–478, 1991.
14. Neufeld, K. L., and White, R. L. Nuclear and cytoplasmic localizations of the adenomatous polyposis coli protein. Proc. Natl. Acad. Sci. USA, *94:* 3034–3039, 1997.
15. Weaver, V. M., Petersen, O. W., Wang, F., Larabell, C. A., Briand, P., Damsky, C., and Bissell, M. J. Reversion of the malignant phenotype of human breast cells in three-dimensional culture and *in vivo* by integrin blocking antibodies. J. Cell Biol., *137:* 231–245, 1997.
16. Campbell, T., Skilton, R. A., Coombes, R. C., Shousha, S., Graham, M. D., and Luqmani, Y. A. Isolation of a lactoferrin cDNA clone and its expression in human breast cancer. Br. J. Cancer, *65:* 19–26, 1992.
17. Close, M. J., Howlett, A. R., Roskelley, C. D., Desprez, P. Y., Bailey, N., Rowning, B., Teng, C. T., Stampfer, M. R., and Yaswen, P. Lactoferrin expression in mammary epithelial cells is mediated by changes in cell shape and actin cytoskeleton. J. Cell Sci., *110:* 2861–2871, 1997.
18. Ronnov-Jessen, L., Petersen, O. W., and Bissell, M. J. Cellular changes involved in conversion of normal to malignant breast: importance of the stromal reaction. Physiol. Rev., *76:* 69–125, 1996.
19. Herwig, S., and Strauss, M. The retinoblastoma protein: a master regulator of cell cycle, differentiation and apoptosis. Eur. J. Biochem., *246:* 581–601, 1997.
20. Gu, W., Schneider, J. W., Condorelli, G., Kaushal, S., Mahdavi, V., and Nadal-Ginard, B. Interaction of myogenic factors and the retinoblastoma protein mediates muscle cell commitment and differentiation. Cell, *72:* 309–324, 1993.
21. Novitch, B. G., Mulligan, G. J., Jacks, T., and Lassar, A. B. Skeletal muscle cells lacking the retinoblastoma protein display defects in muscle gene expression and accumulate in S and G2 phases of the cell cycle. J. Cell Biol., *135:* 441–456, 1996.
22. Noe, V., Alemany, C., Chasin, L. A., and Ciudad, C. J. Retinoblastoma protein associates with SP1 and activates the hamster dihydrofolate reductase promoter. Oncogene, *16:* 1931–1938, 1998.
23. Teng, C. T., Liu, Y., Yang, N., Walmer, D., and Panella, T. Differential molecular mechanism of the estrogen action that regulates *lactoferrin* gene in human and mouse. Mol. Endocrinol., *6:* 1969–1981, 1992.
24. Lussier, M., Filion, M., Compton, J. G., Nadeau, J. H., Lapointe, L., and Royal, A. The mouse keratin 19-encoding gene: sequence, structure and chromosomal assignment. Gene (Amst.), *95:* 203–213, 1990.
25. Berezutskaya, E., Yu, B., Morozov, A., Raychaudhuri, P., and Bagchi, S. Differential regulation of the pocket domains of the retinoblastoma family proteins by the HPV16 E7 oncoprotein. Cell Growth Differ., *8:* 1277–1286, 1997.
26. Mulligan, G., and Jacks, T. The retinoblastoma gene family: cousins with overlapping interests. Trends Genet., *14:* 223–229, 1998.

Bibliography

Mazoyer S, Gayther SA, Nagai MA, Smith SA, Dunning A, van Rensburg EJ, Albertsen H, White R, & Ponder BA. A gene (DLG2) located at 17q12-q21 encodes a new homologue of the Drosophila tumor suppressor dlg-A. Genomics. 1995 Jul 1;28(1):25-31.

Smith SA, Holik PR, Stevens J, Melis R, White R, & Albertsen H. Isolation and mapping of a gene encoding a novel human ADP-ribosylation factor on chromosome 17q12-q21. Genomics. 1995 Jul 1;28(1):113-5.

Brothman AR, Steele MR, Williams BJ, Jones E, Odelberg S, Albertsen HM, Jorde LB, Rohr LR, & Stephenson RA. Loss of chromosome 17 loci in prostate cancer detected by polymerase chain reaction quantitation of allelic markers. Genes Chromosomes Cancer. 1995 Aug;13(4):278-84.

Smith SA, Holik P, Stevens J, Mazoyer S, Melis R, Williams B, White R, & Albertsen H. Isolation of a gene (DLG3) encoding a second member of the discs-large family on chromosome 17q12-q21. Genomics. 1996 Jan 15;31(2):145-50.

Albertsen HM, Smith SA, Melis R, Williams B, Holik P, Stevens J, & White R. Sequence, genomic structure, and chromosomal assignment of human DOC-2. Genomics. 1996 Apr 15;33(2):207-13.

Karpf AR, Peterson PW, Rawlins JT, Dalley BK, Yang Q, Albertsen H, Jones DA. Inhibition of DNA methyltransferase stimulates the expression of signal transducer and activator of transcription 1, 2, and 3 genes in colon tumor cells. Proc Natl Acad Sci U S A. 1999 Nov 23;96(24):14007-12.

Spancake KM, Anderson CB, Weaver VM, Matsunami N, Bissell MJ, & White RL. E7-transduced human breast epithelial cells show partial differentiation in three-dimensional culture. Cancer Res. 1999 Dec 15;59(24):6042-5.

List of Salaried Personnel:

Name:	Position:	Percent contribution to salary:
Hans Albertsen,	Research Instructor,	67%
Jeff Stevens,	Research Associate,	100%
Ray White,	P.I., Professor,	10%